A Global Optimization Algorithm (GOP) For Certain Classes of Nonconvex NLPs : I. Theory

C.A. Floudas* and V. Visweswaran
Department of Chemical Engineering
Princeton University
Princeton, N.J. 08544-5263

April 1990/Revised: July 1990.

Abstract

A large number of nonlinear optimization problems involve bilinear, quadratic and/or polynomial functions in their objective function and/or constraints. In this paper, a theoretical approach is proposed for global optimization in constrained nonconvex **NLP** problems. The original nonconvex problem is decomposed into primal and *relaxed dual* subproblems by introducing new transformation variables if necessary and partitioning of the resulting variable set. The decomposition is designed to provide *valid* upper and lower bounds on the global optimum through the solutions of the primal and *relaxed dual* subproblems respectively. New theoretical results are presented that enable the rigorous solution of the *relaxed dual* problem. The approach is used in the development of a Global OPtimization algorithm **(GOP)**. The algorithm is proved to attain finite ϵ-convergence and ϵ-global optimality. An example problem is used to illustrate the **GOP** algorithm both computationally and geometrically. In an accompanying paper (Visweswaran and Floudas, 1990), application of the theory and the **GOP** algorithm to various classes of optimization problems, as well as computational results of the approach are provided.

Keywords : Global Optimization, Primal-Dual Decomposition, ϵ-optimal solutions.

1 Introduction

Nonlinear programming problems form a major subset of the field of mathematical programming. In particular, problems related to chemical process design and control can often be formulated as nonlinear optimization problems. These problems may involve nonconvexities which imply difficulties with determining a global solution. Chemical engineering examples of such problems abound in reactor network synthesis, phase and chemical reaction equilibrium, heat exchanger network design, batch processes, scheduling and planning, pooling and blending problems, and optimal design of distillation sequences.

For over five decades, there have been a number of algorithms developed for determining local optima for mathematical programming problems. However, the use of a conventional algorithm designed for **NLP** problems is highly dependent on the starting point provided for the algorithm, often leading to the solver failing to determine even a feasible solution. It has been shown that for constrained nonlinear programming problems, the computational complexity of determining whether a given feasible point is a local minimum is an **NP**-complete problem, and the

*Author to whom all correspondence should be addressed.

global optimization of such problems is **NP**-hard (Murty and Kabadi, 1987). For constrained quadratic programming problems, it has also been shown that determining a global solution is **NP**-hard (Pardalos and Schnitger, 1988).

With the advent of advanced computer architectures and the emerging new theoretical results, there has been a growing interest in developing algorithms that locate a global optimum. No attempt is made in this paper to review the different approaches within the global optimization area, since extensive surveys and books on the existing approaches for global optimization are available by Dixon and Szego (1975, 1978), Archetti and Schoen (1984), Pardalos and Rosen (1986, 1987), Torn and Zilinskas (1987), Ratschek and Rokne (1988), Hansen *et al* (1989a,b), Horst and Tuy (1990) and Floudas and Pardalos (1990).

The proposed approaches for global optimization can be largely classified as deterministic or probabilistic approaches. The deterministic approaches include :

(i) Covering methods (e.g. Piyavskii, 1972);

(ii) Branch and bound methods (e.g. Al-Khayyal and Falk, 1983; Horst and Tuy, 1987; Hansen *et al*, 1990);

(iii) Cutting Plane Methods (e.g. Tuy, Thien and Thai, 1985);

(iv) Interval methods (e.g. Hansen, 1979);

(v) Trajectory methods (e.g. Branin, 1972); and

(vi) Penalty methods (e.g. Levy and Montalvo, 1985).

Probabilistic methods for global optimization include :

(i) Random search methods (e.g. Kirkpatrick *et al*, 1983; Pinter, 1984);

(ii) Clustering methods (e.g. Rinnoy Kan and Timmer, 1987); and

(iii) Methods based on statistical models of objective functions (e.g. Zilinskas, 1986).

Stephanopoulos and Westerberg (1975) presented an algorithm for minimizing the sum of separable concave functions subject to linear constraints. The approach utilizes Hestenes' method of multipliers, and accounts for duality gaps in the original problem. Westerberg and Shah (1978) considered structured optimization problems where the problem can be decomposed into several subsystems, and proposed an algorithm for determining whether a given local minimum of such problems is a global solution, through the use of an upper bound on the dual bound for the problem. The algorithm was applied to three heat exchanger network problems involving separable objective functions. However, there is no guarantee of determination of the global solution. Kocis and Grossmann (1988) used Outer Approximation with Equality Relaxation as the basis for a two phase strategy for solving nonconvex **MINLP** algorithms, and implemented the approach in the program DICOPT (Kocis and Grossmann, 1989). These methods solve the original problem through a series of **NLP** (which can be nonconvex) and **MILP** subproblems. This work does not address the global optimum issue of the nonconvex **NLP**, and thus provides no guarantee for global optimality.

Floudas *et al* (1989) utilized the principles of Generalized Benders Decomposition (Geoffrion, 1972) and projected on a subset of variables so as to induce a special convex structure in the primal and master subproblems. Even though global optimality could not be guaranteed, the approach was shown to be very effective for solving nonconvex nonlinear and mixed-integer nonlinear programming problems. The approach was also utilized in developing a global optimum search approach for solving bilinear, definite, indefinite and mixed-integer quadratic programming (Aggarwal and Floudas, 1990), and was applied to the Haverly's pooling problem (Floudas and Aggarwal, 1990).

In this paper, a primal-relaxed dual approach for global optimization is proposed. A statement of the global optimization problem is given in section 2, and a motivating example for this work is presented in section 3. Section 4 presents the duality theory used in the development of the algorithm, while Section 5 contains new theoretical results which form the basis for the global optimization **(GOP)** algorithm. Section 6 describes the global optimization algorithm. Proofs of finite ϵ-convergence and ϵ-global optimality are provided in section 7. The **GOP** algorithm is illustrated through an example problem in section 8, and a geometrical interpretation of the algorithm is given in section 9. Proofs for some of the properties presented in section 5 are detailed in Appendix A, while Appendix B contains three Lemmas used in the proof of finite ϵ-convergence and ϵ-global optimality of the algorithm.

2 Problem Statement

The global optimization problem addressed in this paper is stated as:

Determine a globally ϵ-optimal solution of the following problem:

$$\min_{x,y} f(x,y)$$

$$\begin{aligned} \textit{subject to} \quad g(x,y) &\leq 0 \\ h(x,y) &= 0 \\ x &\in X \\ y &\in Y \end{aligned} \tag{1}$$

where X and Y are non-empty, compact, convex sets, $g(x,y)$ is an m- vector of inequality constraints and $h(x,y)$ is a p-vector of equality constraints. It is assumed that the functions $f(x,y)$, $g(x,y)$ and $h(x,y)$ are continuous, piecewise differentiable and given in analytical form over $X \times Y$. The variables y are defined in such a way that the following *Conditions (A)* are satisfied:

Conditions (A)

(a) $f(x,y)$ is convex in x for every fixed y, and convex in y for every fixed x.

(b) $g(x,y)$ is convex in x for every fixed y, and convex in y for every fixed x.

(c) $h(x,y)$ is affine in x for every fixed y, and affine in y for every fixed x.

At this point, the following question needs to be asked : What are the classes of mathematical problems that can be represented within the framework of (1) and satisfy *Conditions (A)* ? To answer this question, the concepts of *partitioning* and *transformations* have to be introduced.

2.1 Partitioning of the variable set

In some classes of optimization problems, the structure of the problem is such that a direct *partitioning* of the variable set is sufficient to ensure that *Conditions (A)* are satisfied. This is true in the case of quadratic programming problems where the nonconvexities arise solely due to the presence of bilinear terms in the objective function and/or constraints. For such problems, the *projection* of the problem into the space of a subset of the variables results in a convex programming problem. Hence, it is possible to *partition* the variable set into two subsets x and y in such a way that *Conditions (A)* are satisfied (Floudas *et al*, 1989). This is illustrated by the following example.

Illustration:

$$\min_{w} \; -w_1 - w_2$$

$$\begin{aligned} \textit{subject to} \quad w_1 w_2 &\leq 4 \\ 0 \leq w_1 &\leq 4 \\ 0 \leq w_2 &\leq 8 \end{aligned}$$

The nonconvexity in this problem arises due to the presence of the bilinear term $w_1 w_2$ in the first constraint. This naturally suggests *partitioning* the variable set into the subsets w_1 and w_2. Redefining w_1 as x and w_2 as y, it can be seen that for a fixed value of x^k, the constraint is linear in y, and vice-versa. Hence, such a *partition* ensures that *Conditions (A)* are satisfied for this problem.

For larger problems, the choice of subsets for *partitioning* is not so obvious. However, as shown by Floudas *et al* (1989), graph theory can be used to determine the partitions that satisfy *Conditions (A)*.

2.2 Partitioning and Transformations

In the more general case of the nonlinear programming problem, it cannot be expected that simply *partitioning* the variables will ensure that *Conditions (A)* are satisfied (for example, when the nonconvexities are due to general quadratic or polynomial terms). Fortunately, for a number of problems, it is possible to overcome this difficulty by introducing new "transformation" variables so as to reformulate the problem in such a way that the nonconvexities are due to bilinear terms in the objective function and/or constraint set. The resulting variable set can then be partitioned so as to satisfy *Conditions (A)*.

Consider the minimization of the following 6^{th} order polynomial function in one variable subject to bound constraints (Wingo, 1985) :

$$\min_{y} \; \frac{1}{6}y^6 - \frac{52}{25}y^5 + \frac{39}{80}y^4 + \frac{71}{10}y^3 - \frac{79}{20}y^2 - y + \frac{1}{10}$$
$$s.t. \qquad -2 \leq y \leq 11$$

The nonconvexity is due to the terms containing $-y^2$, y^3 and $-y^5$ in the objective function. Since there is only one variable, a simple *partitioning* of the variable set is not possible.

By introducing five new variables x_1 to x_5 and adding the following five inequalities :

$$\begin{aligned} x_1 - y &= 0 \\ x_2 - x_1.y &= 0 \\ x_3 - x_2.y &= 0 \\ x_4 - x_3.y &= 0 \\ x_5 - x_4.y &= 0 \end{aligned}$$

the problem can be converted to a form where the variable set can be partitioned, as shown below:

$$\min \; \frac{1}{6}y^6 - \frac{52}{25}x_5 + \frac{39}{80}y^4 + \frac{71}{10}x_3 - \frac{79}{20}x_2 - y + \frac{1}{10}$$

$$
\begin{aligned}
s.t. \quad x_1 - y &= 0 \\
x_2 - x_1 y &= 0 \\
x_3 - x_2 y &= 0 \\
x_4 - x_3 y &= 0 \\
x_5 - x_4 y &= 0 \\
-2 \le y &\le 11 \\
x_i^L \le x_i &\le x_i^U
\end{aligned}
$$

$$
where \quad x^L = (-2, 0, -8, 0, -32) \quad and \qquad x^U = (11, 121, 1331, 14641, 161051)
$$

For every fixed $y = y^k$, the objective function and the equality constraints are linear, and therefore convex, in x_i, and for every fixed x_i, the objective function is convex in y, and the equality constraints are linear and therefore convex in y. Hence, the reformulated problem satisfies *Conditions (A)*.

Such a reformulation is possible for (a) quadratic programming problems with linear and/or quadratic constraints (including pooling/blending problems), and (b) optimization problems involving polynomial functions of one or more variables in the objective function and/or the constraint set. In general, problems with any combination of bilinear, quadratic or polynomial terms can be made to satisfy *Conditions (A)* by the use of transformation variables and partitions. Thus, the theoretical results of this paper will be applicable to all of the above mentioned classes of nonconvex continuous nonlinear optimization problems.

3 Motivating Example

Consider the following example, which involves the minimization of a nonconvex objective function subject to a linear set of constraints :

$$
\begin{aligned}
&\min_x -x_1 + x_1 x_2 - x_2 \\
s.t. \quad &-6x_1 + 8x_2 \le 3 \\
&3x_1 - x_2 \le 3 \\
&0 \le x_1, x_2 \le 3
\end{aligned}
$$

This problem has a local minimum of -1.0052 at $x = (0.916, 1.062)$, while the global minimum occurs at $x = (1.167, 0.5)$ with an objective value of -1.0833.

When this problem is solved as an **NLP** using **MINOS 5.2** (Murtagh and Saunders, 1988), the solution found is dependent on the starting point. If the solver is provided with the local optimum as a starting point, the solution found is the same point. Perturbation of the variables in a small region around the local optimum shows that the local optimum is a strong local optimum. Hence, a local search technique can fail to find the global optimum at (1.167,0.5).

When the Global Optimum Search (GOS) technique was applied to this problem (Aggarwal and Floudas, 1989) the global optimum was identified from several starting points, but from some of the starting points, the algorithm converged to the local minimum. An explanation of this provided in section 6.3, and it is illustrated in the example problem considered in section 8.

To determine the global solution from the considered initial points, special procedures had to be devised (Aggarwal and Floudas, 1989) that involve properties that exploit the symmetry and utilize a “restart” feature. Further more, regardless of these special schemes, there was no theoretical guarantee that convergence to the global optimum can be obtained from any initial starting point.

It is, however, the very good performance of the Global Optimum Search (Floudas *et al*, 1989) in several classes of problems and the assumptions within it that motivated us towards establishing rigorous theoretical results for both finite ϵ-convergence and ϵ-global optimality that can be applied to different classes of nonconvex nonlinear programming **NLP** problems. Prior to presenting these mathematical properties, the duality theory for problems of form (1) satisfying *Conditions (A)* is presented.

4 Duality Theory

Define the following problem as the **Primal Problem (P)** :

$$\min_{x} \; f(x, y^k)$$

$$\begin{aligned} \textit{subject to} \quad g(x, y^k) &\le 0 \\ h(x, y^k) &= 0 \\ x &\in X \end{aligned} \tag{2}$$

This problem is simply problem (1) solved for fixed values of $y = y^k$. Therefore, it is equivalent to solving the original problem with some additional constraints. Hence, regardless of the value of y^k, any feasible solution of problem (2) provides an upper bound on the solution of problem (1).

The theory for feasible primal problems is first considered in section 4.1 . Section 4.2 considers the case where the primal problem is infeasible for a given fixed value of $y = y^k$.

4.1 Feasible Primal Problems

Problem (1) can also be regarded in an equivalent form where the minimization over the x and y variables occurs separately. That is, problem (1) is equivalent to :

$$\min_{x} \min_{y} f(x, y)$$

$$\begin{aligned} \textit{subject to} \quad g(x, y) &\le 0 \\ h(x, y) &= 0 \\ x &\in X \\ y &\in Y \end{aligned}$$

The *projection* of this problem in the space of the y variables (Geoffrion, 1972) further enables (1) to be written in the following form, featuring inner and outer optimization problems :

$$
\begin{array}{rrcl}
 & \min\limits_{y} & & v(y) \\
\textit{subject to} & v(y) & = & \min\limits_{x \in X} \; f(x,y) \\
 & & & \text{s.t.} \quad h(x,y) = 0 \\
 & & & \qquad\quad g(x,y) \leq 0 \\
 & x & \in & X \\
 & y & \in & Y \cap V \\
\textit{where} & V & \equiv & \{y \; : \; h(x,y) = 0, \; g(x,y) \leq 0 \quad \textit{for some } x \in X\}
\end{array} \tag{3}
$$

Remarks on formulation (3):

(a) The inner minimization problem is parametric in y. For any fixed value of y, say y^k, this problem is simply the primal problem **(P)** solved for that $y = y^k$.

(b) The function $v(y)$ is defined as the set of solutions of (2) for different values of y. Invariably, this is a nonconvex set. Moreover, this set is known only implicitly. For this reason, problem (3) can be very difficult to solve in the form shown above. One way to overcome this difficulty is to consider the inner problem in its dual representation.

From the **Strong Duality Theorem**, if problem (2) satisfies the following conditions:

(a) $f(x,y)$ and $g(x,y)$ are convex in x for every fixed $y = y^k$,

(b) $h(x,y)$ are affine in x for every fixed $y = y^k$,

(c) X is a nonempty, compact convex set, and

(d) For every fixed $y = y^k$, there exists an $\hat{x} \in X$ such that $g(\hat{x}, y^k) < 0$, $h(\hat{x}, y^k) = 0$ and $0 \in \textit{int } h(X)$ where $h(X) \equiv \{h(x) : \; x \in X\}$ (*Constraint Qualification*)

then, the solution of (2), for any fixed $y = y^k$, is identical to the solution of its corresponding dual problem. That is,

$$
\left\{
\begin{array}{rcl}
\min\limits_{x \in X} f(x, y^k) & & \\
\text{s.t.} \quad g(x, y^k) & \leq & 0 \\
h(x, y^k) & = & 0
\end{array}
\right\}
\quad = \quad
\begin{array}{l}
\sup\limits_{\substack{\mu \geq 0 \\ \lambda}} \; \inf\limits_{x \in X} \; \{f(x,y^k) + \lambda^T h(x,y^k) + \mu^T g(x,y^k)\} \\
\qquad \textit{for all } \; y^k \in Y \cap V.
\end{array}
$$

where, λ and μ are the Lagrange multipliers corresponding to the equality and inequality constraints of the primal problem (2).

Remark on the Strong Duality Theorem:

Conditions (A) ensure that conditions (a) and (b) of the Strong Duality Theorem are satisfied. Condition (c) is also satisfied by the definition of the problem (1). Therefore, the Strong Duality Theorem will hold for this problem as long as the *constraint qualification* (condition (d)) is satisfied.

The use of the Strong Duality Theorem permits the set $v(y)$ to be written as

$$v(y) = \sup_{\substack{\mu \geq 0 \\ \lambda}} \inf_{x \in X} \{f(x,y) + \lambda^T h(x,y) + \mu^T g(x,y)\} \quad \forall\, y \in Y \cap V$$

From the definition of supremum, the maximization over λ and μ can be relaxed to an upper bound :

$$v(y) \geq \inf_{x \in X} \{\ f(x,y) + \lambda^T h(x,y) + \mu^T g(x,y)\ \} \quad \textit{for all} \quad \mu \geq 0, \lambda.$$

Thus, assuming the existence of feasible solutions to the inner minimization problems, the dual representation of $v(y)$ leads to the following formulation, equivalent to (3) :

$$\min_y \ v(y)$$

subject to

$$\begin{aligned} v(y) &\geq \min_{x \in X}\{\ f(x,y) + \lambda^T h(x,y) + \mu^T g(x,y)\ \}, \quad \forall \mu \geq 0, \lambda \\ y &\in Y \cap V \\ V &\equiv \{y : h(x,y) = 0, g(x,y) \leq 0 \textit{ for some } x \in X\} \end{aligned} \tag{4}$$

Remark on formulation (4) :

The last two sets of constraints in problem (4) represent an implicitly defined set in y. Therefore, the presence of these constraints makes the solution of (4) extremely difficult. This can be avoided by simply dropping the two constraints from (4). This is equivalent to relaxing the constraint region for the problem, and thus represents a lower bound for the solution of the problem.

By dropping the last two constraints from (4), the **Relaxed Dual (RD)** is obtained:

$$\min_{\substack{y \in Y \\ \mu_B}} \ \mu_B$$

$$\textit{subject to} \quad \mu_B \geq \min_{x \in X}\{f(x,y) + \lambda^T h(x,y) + \mu^T g(x,y)\}, \quad \forall \mu \geq 0, \lambda \tag{5}$$

where μ_B is a scalar.

The inner minimization problem is denoted as the **Inner Relaxed Dual (IRD)** problem. This problem is:

$$\min_{x \in X} \ L(x, y, \lambda^k, \mu^k) \tag{6}$$

where

$$L(x, y, \lambda^k, \mu^k) = f(x,y) + {\lambda^k}^T h(x,y) + {\mu^k}^T g(x,y) \tag{7}$$

is the Lagrange function formulated for problem (2) at the kth iteration.

Remarks on the Relaxed Dual (RD) Problem :

1. The use of projection and duality theory as presented provides an obvious way of obtaining upper and lower bounds on the global solution of (1). The primal problem, given by (2), represents an upper bound on the original problem (1). The relaxed dual **(RD)** problem in the form (5) contains fewer constraints than (4). Hence, it represents a valid lower bound on the original problem (1). Thus, these two problems can be used in an iterative fashion to determine a global solution of (1).

2. In the form given by (5), the relaxed dual **(RD)** problem is difficult to solve, since it contains the inner relaxed dual **(IRD)** problem, which is parametric in y (infinite programming problem in y). Hence, it is necessary to determine an equivalent problem that has only constraints and still provides a valid underestimator of the relaxed dual problem, enabling it to be solved rigorously. Section 5 presents theoretical properties, by the use of which this can be achieved.

4.2 Infeasible Primal Subproblems

In cases where the primal subproblem (2) is infeasible, a third subproblem must be solved for generating the appropriate values of λ and μ. One possible formulation for this problem is

$$\min_{\substack{\beta^+,\beta^-,\alpha\geq 0\\ x\in X}} \quad \sum_{i=1}^{m}\alpha_i + \sum_{i=1}^{p}(\beta_i^+ + \beta_i^-)$$

$$\begin{aligned} h(x,y)+\beta^+-\beta^- &= 0 \\ g(x,y)-\alpha &\leq 0 \end{aligned} \tag{8}$$

In problem (8), for every fixed y^k, the objective function is linear, the equality constraints are linear and the inequality constraints are convex. If $\delta = \sum_{i=1}^{m}\alpha_i + \sum_{i=1}^{p}(\beta_i^+ + \beta_i^-)$, then, the strong duality theorem provides

$$\left\{ \begin{array}{rcl} & \min\delta & \\ s.t.\ \ g(x,y)-\alpha & \leq & 0 \\ h(x,y)+\beta^+-\beta^- & = & 0 \end{array} \right\} = \max_{\substack{\mu_1\geq 0\\ \lambda_1}} \min_{x\in X} \quad \begin{array}{l} \{\delta + \lambda_1^T(h(x,y)+\beta^+-\beta^-) \\ \quad +\mu_1^T(g(x,y)-\alpha) \quad \} \end{array} \tag{9}$$

where λ_1 and μ_1 are the Lagrange multipliers for the equality and inequality constraints for the solution of (8) for fixed $y = y^k$. If $\overline{\delta}$ is the optimal solution of (8), then (9), together with the optimality conditions for problem (8), implies that

$$\overline{\delta} = \max_{\substack{\lambda_1\\ \mu_1\geq 0}} \min_{x\in X}\{\lambda_1^T h(x,y) + \mu_1^T g(x,y)\}$$

Since we seek to minimize the infeasibilities $\overline{\delta}$, this can be used as a constraint for the master problem in the following form :

$$\max_{\substack{\lambda_1\\ \mu_1\geq 0}} \min_{x\in X}\{\lambda_1^T h(x,y) + \mu_1^T g(x,y)\} = 0$$

A relaxed form of this constraint is

$$\min_{x\in X} L_1(x,y,\lambda_1,\mu_1) \leq 0 \tag{10}$$

where

$$L_1(x,y,\lambda_1,\mu_1) = \lambda_1^T h(x,y) + \mu_1^T g(x,y) \tag{11}$$

This is the appropriate constraint to be added to the master problem for those iterations where the primal problem is infeasible.

5 Mathematical Properties

The mathematical properties are presented first in section 5.1 for the case of feasible primal subproblems. The equivalent properties for iterations involving infeasible primal subproblems are considered in section 5.2 . Insights on the mathematical properties are provided through the illustration presented in Section 2.1 .

5.1 Feasible Primal Subproblems

Property 1 : *The solution of each primal subproblem* **(P)** *is the global solution of that problem.*

Proof: From the criteria given by *Conditions (A)*, the primal problem is a convex programming problem in x, and hence its global optimality conditions (Avriel, 1976) are satisfied. Hence, the solution of every primal subproblem is its global solution . □

Illustration - Continued: As detailed in section 2.1, by selecting w_1 as x and w_2 as y, the variable set can be partitioned so as to satisfy *Conditions (A)*. For a fixed value of $y = y^k$, the primal subproblem for this example can be written as:

$$\min \ -x - y^k$$

$$\begin{aligned} s.t. \quad xy^k &\leq 4 \\ 0 - x &\leq 0 \\ x - 4 &\leq 0 \end{aligned}$$

This is a linear programming problem in x. Hence, property 1 holds.

Property 2 : *The Lagrange function of the k th iteration,* $L(x, y, \lambda^k, \mu^k)$*, satisfies the following Conditions (B) :*

Conditions (B):

(i) $L(x, y, \lambda^k, \mu^k)$ *is convex in* $x \quad \forall \ y = y^d$*, and*

(ii) $L(x, y, \lambda^k, \mu^k)$ *is convex in* $y \quad \forall \ x = x^k$*.*

Proof : The Lagrange function is a weighted sum of the objective function, equality and inequality constraints of the primal subproblem. For each discretized $y = y^d$, the Lagrange function is convex in x since it is a sum of convex functions in x (by *Conditions (A)*). Similarly, for every fixed $x = x^k$, the Lagrange function is a sum of convex functions in y (by *Conditions (A)*). □

Illustration - Continued :

$$L(x, y, \mu^k) = -x - y + \mu_1^k(xy - 4) + \mu_2^k(0 - x) + \mu_3^k(x - 4)$$

As can be seen, once either x or y is fixed, the Lagrange function is linear, and thus convex, in the other variable.

Property 3 : *The solution of the* **Inner Relaxed Dual (IRD)** *problem is its global solution for each discretized* $y = y^d$.

Proof : From Property 2, the Lagrange function is convex in x for every fixed (discretized) $y = y^d$. Therefore, the Inner Relaxed Dual **(IRD)** problem also meets the global optimality requirements (Avriel, 1976). □

Property 4 :

$$\min_{x \in X} L(x, y^d, \lambda^k, \mu^k) \geq \min_{x \in X} L(x, y^d, \lambda^k, \mu^k)\big|_{x^k}^{lin} \qquad \forall y = y^d \tag{12}$$

where $L(x, y^d, \lambda^k, \mu^k)\big|_{x^k}^{lin}$ *is the linearization of the Lagrange function* $L(x, y^d, \lambda^k, \mu^k)$ *around* x^k*, the solution of the* k^{th} *primal subproblem.*

Proof : For every discretized $y = y^d$, the Lagrange function $L(x, y^d, \lambda^k, \mu^k)$ is convex in x. Hence, the linearization of the Lagrange function around some x is a valid underestimator of the function for all values of x. Since this is true regardless of the point of linearization of x,

$$L(x, y^d, \lambda^k, \mu^k) \geq L(x, y^d, \lambda^k, \mu^k)\big|_{x^k}^{lin} \qquad \forall y = y^d \tag{13}$$

Since this applies for every x, the minimum over x can be taken on both sides leading to the desired result. □

Remark : For every fixed y, the linearization of the Lagrange function around x^k is given by

$$\begin{aligned} L(x, y, \lambda^k, \mu^k)\big|_{x^k}^{lin} &= L(x^k, y, \lambda^k, \mu^k) + \nabla_x L(x, y, \lambda^k, \mu^k)\big|_{x^k}.(x - x^k) \\ &= L(x^k, y, \lambda^k, \mu^k) + \sum_{i=1}^{n} \nabla_{x_i} L(x, y, \lambda^k, \mu^k)\big|_{x^k}.(x_i - x_i^k) \end{aligned}$$

where x_i is the ith x variable, $i = 1, 2...n$. From this, it can be seen that $L(x, y, \lambda^k, \mu^k)\big|_{x^k}^{lin}$ can be written as:

$$L(x, y, \lambda^k, \mu^k)\big|_{x^k}^{lin} = \Phi_1(y, \lambda^k, \mu^k).\Psi_1(x) + \Phi_2(y, \lambda^k, \mu^k) \tag{14}$$

where

$$\Phi_1(y, \lambda^k, \mu^k) = \nabla_x L(x, y, \lambda^k, \mu^k)\big|_{x^k} \tag{15}$$

$$\Psi_1(x) = x - x^k \tag{16}$$

$$\Phi_2(y, \lambda^k, \mu^k) = L(x^k, y, \lambda^k, \mu^k) \tag{17}$$

In general, if a function $\Psi_2(x, \lambda^k, \mu^k)$ is used to represent terms in the Lagrange function or its linearization that are linear in the x variables and independent of the y variables, then

(i) If the Lagrange function is linear in x $\forall\, y = y^d$ and linear in y $\forall\, x = x^k$, then $L(x, y, \lambda^k, \mu^k)$ can be written as:

$$L(x, y, \lambda^k, \mu^k) = \Phi_1(y, \lambda^k, \mu^k).\Psi_1(x) + \Psi_2(x, \lambda^k, \mu^k) + \Phi_2(y, \lambda^k, \mu^k) \tag{18}$$

(ii) If the Lagrange function is convex in x $\forall\, y = y^d$ and convex in y $\forall\, x = x^k$, then the linearization of $L(x, y, \lambda^k, \mu^k)$ around x^k can be written as

$$L(x, y, \lambda^k, \mu^k)\big|_{x^k}^{lin} = \Phi_1(y, \lambda^k, \mu^k).\Psi_1(x) + \Psi_2(x, \lambda^k, \mu^k) + \Phi_2(y, \lambda^k, \mu^k) \tag{19}$$

Based upon this form of the Lagrange function, the following definition is made:

Definition : The subset of variables x that exist in the function $\Psi_1(x)$ are called the *connected* variables.

Illustration - Continued:

$$\begin{aligned} \Phi_1(y, \lambda^k, \mu^k) &= \mu_1^k y \\ \Psi_1(x) &= x \\ \Phi_2(y, \lambda^k, \mu^k) &= -y - 4\mu_1^k - 4\mu_3^k \\ \Psi_2(x, \lambda^k, \mu^k) &= -(1 + \mu_2^k - \mu_3^k)x \end{aligned}$$

Thus, for this problem, x is a *connected* variable.

Property 5 : *The optimal solution of the inner relaxed dual* **(IRD)** *problem, with the Lagrange function replaced by its linearization around* x^k*, depends only on those* x_i*, for which* $\Psi_1(x)$ *is a function of* x_i (i.e. the *connected* variables.)

Proof : See Appendix A.

Illustration of Property 5 :

This property can be illustrated by the following example involving two x and one y variable.

$$\min_x x_1 y_1 + x_2$$

$$\begin{aligned} s.t. \quad -x_1 + 4y_1 - 2 &\leq 0 \\ 3x_1 - x_2 - 3 &\leq 0 \end{aligned}$$

The solution of the primal problem for this example, for a fixed value of $y_1 = 1$ provides $x_1 = 2$, $x_2 = 3$, and the Lagrange multipliers $\mu_1 = 4$ and $\mu_2 = 1$. From this solution, the Lagrange function is given as

$$\begin{aligned} L(x, y, \mu) &= x_1 y_1 + x_2 + 4(-x_1 + 4y_1 - 2) + 1(3x_1 - x_2 - 3) \\ &= x_1(y_1 - 1) + 16y_1 - 11 \end{aligned}$$

Thus, it can be seen that when the Lagrange function is formulated with the Lagrange multipliers from the solution of the primal problem, the terms in x_2 cancel out. This will occur for every variable such as x_2 which is not bilinearly involved with any y variable, that is, for any x which is not a *connected* variable. The same can be proved when a variable such as x_2 occurs in a convex fashion in the problem, but is not *connected.* In such a case, the terms involving that variable will not vanish in the Lagrange function itself, but will do so in the linearization of the Lagrange function around the solution of the corresponding primal problem.

Remark: This property is important from the computational point of view. It implies that the inner relaxed dual **(IRD)** problem could be replaced by a problem involving the minimization of the linearization of the Lagrange function over the set of *connected* x variables. For a number of optimization problems, the nonconvexities are limited to a small number of terms and variables. This property suggests that for such problems, the computational effort required to obtain a global solution is not determined by the number of variables in the problem, but rather by the number of *connected* variables, which can result in reductions by several orders of magnitude in the time taken to solve the problem.

Property 6: *Define* I_c *to be the set of connected* x *variables,* B_j *to be a combination of LOWER/UPPER bounds of these variables and CB to be the set of all such combinations. Then, if the Lagrange function is linear in* x $\forall y = y^d$,

$$\min_{x \in X} L(x, y, \lambda^k, \mu^k) \geq \min_{B_j \in CB} L(x^{B_j}, y, \lambda^k, \mu^k) \quad \forall y \tag{20}$$

Proof: See Appendix A.

Illustration - Continued: It can be seen that the Lagrange function, which is given by

$$L(x, y, \mu_1^k, \mu_2^k, \mu_3^k) = -x - y + \mu_1^k(xy - 4) + \mu_2^k(0 - x) + \mu_3^k(x - 4)$$

is linear in x for every fixed $y = y^d$. Hence, for this problem, the solution of the inner relaxed dual **(IRD)** problem, for any value of y, will lie at a bound of x. There are two possible bounds for x, its upper and lower bound.

Remark: Property 6 suggests that it is sufficient to set the *connected* variables x_i , $i \in I_c$ to a set of bounds B_j, and solving the Relaxed Dual problem once for every such combination of bounds. However, no information is provided about which Lagrange functions from previous iterations could be included in the Relaxed Dual problem. This is provided by the following property.

Property 7: *Suppose that the optimal solution of the Inner Relaxed Dual* **(IRD)** *problem occurs at* $\overline{x}$ *; that is, for every* $y \in Y$*, there exists an* $\overline{x} \in X$ *satisfying*

$$L^*(\overline{x}, y, \lambda^k, \mu^k) = \min_{x \in X} L(x, y, \lambda^k, \mu^k)$$

Then, for every k,

$$\min_{x \in X} L(x, y, \lambda^k, \mu^k) \geq \min_{B_j \in CB} \left\{ \begin{array}{ll} L(x^{B_j}, y, \lambda^k, \mu^k)\big|_{x^k}^{lin} & \\ with \quad \nabla_{x_i} L(x, y, \lambda^k, \mu^k)\big|_{x^k} \geq 0 & \forall x_i^{B_j} = x_i^L \ , \\ \quad\quad\quad \nabla_{x_i} L(x, y, \lambda^k, \mu^k)\big|_{x^k} \leq 0 & \forall x_i^{B_j} = x_i^U \end{array} \right\} \forall y. \tag{21}$$

where x_i^L *and* x_i^U *are the LOWER and UPPER bounds on* connected x_i *respectively.*

Proof : See Appendix A.

Definition: The constraints requiring the positivity or negativity of the gradients of a particular Lagrange function w.r.t x_i are called the *qualifying constraints* of that Lagrange function.

Remark: For each discretized $y = y^d$, the nature of the actual bound of x where the solution lies depends on the sign of the gradient of the Lagrange function. As a consequence of this property, it is sufficient to evaluate the Lagrange function over all the possible combinations of the bounds for the *connected* x_i, with the corresponding constraints of the gradients of $L(x, y, \lambda^k, \mu^k)$ with respect to the *connected* x variables.

Illustration - Continued : The KKT gradient condition for the kth primal problem is given by

$$\nabla_x L(x, y^k, \mu^k) = -1 + \mu_1^k y^k - \mu_2^k + \mu_3^k = 0$$

so that

$$1 + \mu_2^k - \mu_3^k = \mu_1^k y^k$$

Using this condition, the Lagrange function for the relaxed dual **(RD)** problem can be written as

$$\begin{aligned} L(x, y, \mu^k) &= -x - y + \mu_1^k(xy - 4) + \mu_2^k(0 - x) + \mu_3^k(x - 4) \\ &= (-1 - \mu_2^k + \mu_3^k + \mu_1^k y)x - y - 4\mu_1^k - 4\mu_3^k \\ &= \mu_1^k(y - y^k)x - y - 4\mu_1^k - 4\mu_3^k \end{aligned}$$

Thus, the gradient of the Lagrange function *w.r.t* x is

$$\nabla_x L(x, y, \mu^k) = \mu_1^k(y - y^k)$$

The relaxed dual **(RD)** problem is solved for the two bounds of x. In each case, the domain of y is restricted by the corresponding *qualifying* constraint. This constraint is given by :

$$\begin{array}{lll} If \quad x^B = 0, & then \quad \mu_1^k(y - y^k) \ \geq \ 0 & and \\ If \quad x^B = 4, & then \quad \mu_1^k(y - y^k) \ \leq \ 0 & \end{array}$$

In other words, the relaxed dual problem needs to be solved twice, once for $y \leq y^k$ with x set to its *upper* bound ($x^B = x^U = 4$), and once for $y \geq y^k$ with x set to its *lower* bound ($x^B = x^L = 0$).

Property 8 : *If $\Phi_1(y, \lambda^k, \mu^k)$ is linear in y, then $\nabla_x L(x, y, \lambda^k, \mu^k)\big|_{x^k}$, forms a linear set of constraints in y.*

Proof : See Appendix A.

Remark :

This assumption of $\Phi_1(y, \lambda^k, \mu^k)$ being linear in y is valid for several classes of problems including those mentioned in section 2, namely, for problems involving nonconvexities due to bilinear, quadratic, or polynomial terms in either the objective function or constraints.

Illustration - Continued : The assumption holds, since

$$\Phi_1(y, \lambda^k, \mu^k) = \mu_1^k . y$$

is linear in y. It can be easily seen that :

$$\nabla_x L(x, y, \lambda^k, \mu^k)\big|_{x^k} = \mu_1^k(y - y^k)$$

which is a linear constraint in y.

Remark: Property 8 ensures that the *qualifying* constraint for every Lagrange function can be introduced into the Relaxed Dual problem without destroying the convexity of the problem.

Property 9 :*At the Kth iteration, define*

(i) *UL(k,K) to be the set of Lagrange functions from the kth iteration ($k < K$) whose qualifying constraints are satisfied at y^K, the fixed value of the complicating variables for the Kth primal subproblem.*

(ii) $\overline{\mu}_B{}^K$ *to be the optimal value of the K th Relaxed Dual Problem. That is,*

$$\overline{\mu}_B{}^K = \left\{ \begin{array}{ll} & \min\limits_{y, \mu_B} \ \mu_B \\ s.t. & \mu_B \geq \min\limits_{x \in X} L(x, y, \lambda^k, \mu^k) \quad k = 1, 2....(K-1) \\ & \mu_B \geq \min\limits_{x \in X} L(x, y, \lambda^K, \mu^K) \end{array} \right\}$$

Then,

$$\overline{\mu}_B{}^K \geq \min_{B_l \in CB} \left\{ \begin{array}{l} \min\limits_{y,\mu_B} \ \mu_B \\ \\ s.t. \\ \quad \left. \begin{array}{ll} \mu_B \geq & L(x^{B_j}, y, \lambda^k, \mu^k)\big|_{x^k}^{lin} \\ & \nabla_{x_i} L(x, y, \lambda^k, \mu^k)\big|_{x^k} \leq 0 \quad if \ x_i^{B_j} = x_i^U \\ & \nabla_{x_i} L(x, y, \lambda^k, \mu^k)\big|_{x^k} \geq 0 \quad if \ x_i^{B_j} = x_i^L \end{array} \right\} \begin{array}{l} \forall j \in UL(k, K) \\ k = 1, 2....K-1 \end{array} \\ \\ \quad \begin{array}{ll} \mu_B \geq & L(x^{B_l}, y, \lambda^K, \mu^K)\big|_{x^k}^{lin} \\ & \nabla_{x_i} L(x, y, \lambda^K, \mu^K)\big|_{x^K} \leq 0 \quad if \ x_i^{B_l} = x_i^U \\ & \nabla_{x_i} L(x, y, \lambda^K, \mu^K)\big|_{x^K} \geq 0 \quad if \ x_i^{B_l} = x_i^L \end{array} \end{array} \right\} \tag{22}$$

Proof : See Appendix A.

Remark: Property 9 is the final link in the series of rigorous simplifications made in the inner Relaxed Dual problem. From properties 4-9, the Relaxed Dual problem can be replaced by the RHS of (22).

Illustration - Continued: To determine the set of Lagrange functions from the previous iterations that should be present in the current relaxed dual problems, the *qualifying* constraints for those Lagrangians are evaluated at y^K, the fixed value of y that was used in the current primal problem. If the constraint is satisfied, then that Lagrange function, along with the corresponding *qualifying* constraint, is added to the current relaxed dual problems. This implies that from each of the previous iterations, only one Lagrange function can be present. Thus, for example, from the kth iteration, the Lagrange function with $x^B = 0$ would have the following qualifying constraint:

$$y - y^k \geq 0$$

If this constraint is satisfied at y^K, i.e. if $y^K - y^k \geq 0$, then the Lagrange function along with the constraint $y - y^k \geq 0$ will be present in each of the two relaxed dual problems for the current (Kth) iteration.

Suppose that the algorithm is at the 2^{nd} iteration, with $y^2 = 8$, and let $y^1 = 2$. There are two Lagrange functions to be selected from the first iteration. The *qualifying* constraints for these Lagrange functions are evaluated at $y = y^2 = 8$. This leads to :

(i) For the Lagrange function from the first iteration with x set to the lower bound ($x^B = x^L = 0$), the *qualifying* constraint is

$$y - y^1 \geq 0$$

Evaluation of this constraint at $y = y^2 = 8$ gives

$$y^2 - y^1 = 8 - 2 = 6 \geq 0$$

Thus, this *qualifying* constraint is satisfied at $y = y^2 = 8$. This implies that the corresponding Lagrange function should be present for the current relaxed dual problems.

(ii) For the Lagrange function from the first iteration with x set to the upper bound ($x^B = x^U = 4$), the *qualifying* constraint is

$$y - y^1 \leq 0$$

Evaluation of this constraint at $y = y^2 = 8$ gives

$$y^2 - y^1 = 8 - 2 = 6 \not\leq 0$$

Since this *qualifying* constraint is *not* satisfied at $y = y^2 = 8$, the corresponding Lagrange function cannot be present in the current relaxed dual problems.

Thus, at the second iteration, the following Lagrange function and its corresponding *qualifying* constraint from the first iteration is present:

$$\begin{aligned} \mu_B &\geq L^1(x = 0, y, \mu_1^1, \mu_2^1, \mu_3^1) \\ & y - 2 \geq 0 \end{aligned}$$

where L^1 is the Lagrange function from the first iteration.

At the second iteration, there are two relaxed dual **(RD)** problems to be solved - one for $y \geq 8$, and once for $y \leq 8$. The relaxed dual problem for $y \geq 8$, with the Lagrange function from the second iteration evaluated at $x^B = x^L = 0$, is shown below :

$$\begin{aligned} & \min_{y, \mu_B} \ \mu_B \\ s.t. \quad \mu_B &\geq L^1(x = 0, y, \mu_1^1, \mu_2^1, \mu_3^1) \\ & y - 2 \geq 0 \\ \mu_B &\geq L^2(x = 0, y, \mu_1^2, \mu_2^2, \mu_3^2) \\ & y - 8 \geq 0 \end{aligned}$$

Similarly, the other relaxed dual problem for the 2^{nd} iteration, that is solved for the region $y - 8 \leq 0$ (so that $x^B = x^U = 4$), can be written as :

$$\begin{aligned} & \min_{y, \mu_B} \ \mu_B \\ s.t. & \\ \mu_B &\geq L^1(x = 0, y, \mu_1^1, \mu_2^1, \mu_3^1) \\ & y - 2 \geq 0 \\ \mu_B &\geq L^2(x = 4, y, \mu_1^2, \mu_2^2, \mu_3^2) \\ & y - 8 \leq 0 \end{aligned}$$

The only difference in the formulation of these two problems is in the last two constraints, that is, the constraints that are formulated from the solution of the current (2^{nd}) primal problem. This is because once the right Lagrange function and its accompanying *qualifying* constraint from the first iteration has been selected to be present in the current relaxed dual problems, it has the x set to the appropriate bound, (in this case, $x^B = x^L = 0$) and therefore is a function only of y. Consequently, this constraint remains the same for both the relaxed dual problems solved at the 2^{nd} iteration.

Property 10 : *The solution of each relaxed dual subproblem in the form given by the RHS of (22) is its global solution*

Proof: At the Kth iteration, the Relaxed Dual **(RD)** problem is given by the RHS of (22). The Lagrange functions as used in the Relaxed Dual problem are functions only of y. From *Conditions (B)*, they are convex functions of y.

The gradients of the Lagrange functions w.r.t x_i are linear in y and therefore convex in y. Since the Relaxed Dual (**RD**) problem involves the minimization of a scalar subject to a number of convex inequality constraints in y, it satisfies the global optimality conditions (Avriel, 1976). □

5.2 Infeasible Primal Subproblems

In cases where the primal subproblem (2) is infeasible, the Lagrange function to be added to the relaxed dual (**RD**) problem is of the form given by (10) and (11). This Lagrange function is similar to the form of the Lagrange function for the iterations when the primal problem is feasible. It can be easily shown that the Properties 2-10 presented in subsection 5.1 are all applicable directly for the case of infeasible primal problems by simply replacing μ_B by 0 and $L(x, y, \lambda^k, \mu^k)$ by $L_1(x, y, \lambda^k, \mu^k)$. The constraints to be added along with the Lagrange function to the Relaxed Dual problem are again of the form

$$\begin{aligned} \nabla_{x_i} L_1(x, y, \lambda^k, \mu^k)|_{x^k} &\leq 0 \quad or \\ \nabla_{x_i} L_1(x, y, \lambda^k, \mu^k)|_{x^k} &\geq 0 \end{aligned}$$

depending on whether the variable x_i is at its upper or its lower bound respectively.

Remark: It should be noted that for unconstrained optimization problems or for problems where it is possible to incorporate the constraint set in the relaxed dual problem (e.g. quadratic problems where transformation variables are used for every x, and concave minimization), there will be no infeasible primal problems. This is also true for problems involving only one joint constraint in x and y such as the illustrating example, where the constraint can be used to generate a bound for the y variable, thus ensuring that it is always feasible *w.r.t* the primal problem.

6 The Global OPtimization (GOP) Algorithm

Based on the properties presented in the previous section, it is possible to solve problem (1) through the use of upper and lower bounds obtained by solving a primal problem of the form (2) and a relaxed dual problem where the inner optimization problem has been replaced by the right hand side of (22). In this form, the relaxed dual problem contains only constraints and not inner optimization problems, and therefore it is possible to solve both the primal and relaxed dual problems using a nonlinear programming solver. Thus, it is possible to iterate between these two problems until the two bounds are within ϵ. This iterative procedure has been used in developing the **GOP** algorithm. A schematic of the algorithm is shown in Figures 1(a) and 1(b).

First, an initial set of values y^1 is selected for the complicating variables. A counter K is set to 1, and sets K^{feas} and K^{infeas} are set to empty sets. Next, the primal problem (**P**) of the form shown in section 4 is solved for this fixed value of y. If the primal problem is feasible, the set K^{feas} is updated to contain K, and the solution x^K and the optimal Lagrange multipliers λ^K and μ^K obtained from the primal problem are stored. The upper bound is updated, being set equal to the lowest feasible solution of the primal problems obtained up to the current iteration. If the primal problem is infeasible, the set K^{infeas} is updated to contain K, and a relaxed primal problem is solved. The optimal multiplier vectors λ_1^K and μ_1^K are again stored.

Before solving the relaxed dual problem, the Lagrange functions from all the previous iterations that can be used as constraints for the current relaxed dual problem are determined. To achieve this, the qualifying constraints of every such Lagrange function are evaluated at y^K, the fixed value of the complicating variables for the Kth (current) primal problem. If the qualifying constraints are satisfied at y^K, then the Lagrangian and its accompanying qualifying constraints are selected to be constraints for the current relaxed dual problem. This criterion is applied for *every* previous iteration, even if the primal problem is infeasible for that iteration.

The relaxed dual problem is then solved for each combination of x^{B_j} . In each case, the Lagrange function formulated from the *current* primal problem is chosen as a constraint for the relaxed dual problem, with the Lagrangian being used with $x = x^{B_j}$. In addition to this, a corresponding constraint for the gradient of the Lagrange function is added to the relaxed dual problem. The optimal values of μ_B for every such relaxed dual problem, and the corresponding optimal y, are stored.

After the relaxed dual problem has been solved for every possible combination of the bounds x^{B_j} , the only remaining task is to determine a new lower bound for the problem and select a fixed value of y for the next primal problem. This is done by taking the minimum of all the stored values of μ_B as the lower bound, and the corresponding stored values of y as the y^{K+1} for the next primal. Once a particular μ_B and y have been selected, they are dropped from the stored set. This is to ensure that the relaxed dual problem will not return the same value of y and μ_B during successive problems, except during the final convergence stage of the algorithm, when the solutions of the relaxed dual problems at successive iterations can be very close to each other.

Finally, the check for convergence is done. If the lower bound from the relaxed dual problem is within ϵ of the upper bound from the primal problems, an ϵ-optimal solution to the original problem has been reached, and the algorithm can be stopped. Else, the algorithm continues with the solution of more primal and relaxed dual problems.

6.1 The GOP algorithm

The **GOP** algorithm can be formally stated in the following steps:

STEP 0– *Initialization of parameters:*

Define the storage parameters $\mu_B^{stor}(K^{max}, B_j)$, $y^{stor}(K^{max}, B_j)$ and $y^k(K^{max}, B_j)$ over the set of bounds CB and the maximum expected number of iterations K^{max}. Define P^{UBD} and M^{LBD} as the upper and lower bounds obtained from the primal and relaxed dual problems respectively. Set

$$\mu_B^{stor}(K^{max}, B_j) = U,\ \ P^{UBD} = U,\ \ \textit{and}\ M^{LBD} = L.$$

where U is a very large *positive* number and L is a very large *negative* number. Select an initial set of values y^1 for the complicating variables. Set the counter K equal to 1, and sets K^{feas} and K^{infeas} to empty sets. Select a convergence tolerance parameter ϵ. Identify the set of *connected* x variables I_c.

STEP 1– *Primal problem:*

Store the value of y^K. Solve the following primal problem :

$$\min_{x \in X} \ f(x, y^K)$$

$$\begin{aligned} \textit{subject to} \quad g(x, y^K) &\leq 0 \\ h(x, y^K) &= 0 \end{aligned}$$

If the primal problem is feasible, update the set K^{feas} to contain K, and store the optimal Lagrange multipliers λ^K and μ^K. Update the upper bound so that

$$P^{UBD} = MIN(P^{UBD}, P^K(y^K))$$

where $P^K(y^K)$ is the solution of the current (K th) primal problem. If the primal problem is infeasible, update the set K^{infeas} to contain K, and solve the following relaxed primal subproblem :

$$\min_{\substack{\alpha_i, \beta_i^+, \beta_i^- \geq 0 \\ x \in X}} \sum_i \alpha_i + \sum_i (\beta_i^+ + \beta_i^-)$$

$$\begin{aligned} \textit{subject to} \quad g(x, y^K) - \alpha &\leq 0 \\ h(x, y^K) + \beta^+ - \beta^- &= 0 \end{aligned}$$

Store the values of the optimal Lagrange multipliers λ_1^K and μ_1^K.

STEP 2- *Selection of Lagrangians from the previous iterations:*

For $k = 1, 2..K-1$, evaluate all the *qualifying* constraints of every Lagrange function (i.e., corresponding to each set of bounds of x) at y^K. If every *qualifying* constraint of a Lagrange function is satisfied at y^K, select that Lagrange function to be in the set $UL(k, K)$, that is, to be present in the current relaxed dual problems along with its *qualifying* constraints.

STEP 3– *Relaxed Dual Problem:*

Formulate the Lagrange function corresponding to the current primal problem. Add this as a constraint to the relaxed dual problem. Then:

(a) Select a combination of the bounds of the *connected* variables in x, say $B_l = B_1$.

(b) Solve the following relaxed dual problem:

$$\min_{y \in Y, \mu_B} \mu_B$$

subject to

$$\left.\begin{aligned} \mu_B \geq \quad & L(x^{B_j}, y, \lambda^k, \mu^k)\big|_{x^k}^{lin} \\ & \nabla_{x_i} L(x, y, \lambda^k, \mu^k)\big|_{x^k} \leq 0 \quad if\ x_i^{B_j} = x_i^U \\ & \nabla_{x_i} L(x, y, \lambda^k, \mu^k)\big|_{x^k} \geq 0 \quad if\ x_i^{B_j} = x_i^L \end{aligned}\right\} \quad \begin{array}{c} \forall j \in UL(k, K) \\ k = 1, 2 \ldots K-1 \\ k \in K^{feas} \end{array}$$

$$\left.\begin{aligned} 0 \geq \quad & L_1(x^{B_j}, y, \lambda_1^k, \mu_1^k)\big|_{x^k}^{lin} \\ & \nabla_{x_i} L_1(x, y, \lambda_1^k, \mu_1^k)\big|_{x^k} \leq 0 \quad if\ x_i^{B_j} = x_i^U \\ & \nabla_{x_i} L_1(x, y, \lambda_1^k, \mu_1^k)\big|_{x^k} \geq 0 \quad if\ x_i^{B_j} = x_i^L \end{aligned}\right\} \quad \begin{array}{c} \forall j \in UL(k, K) \\ k = 1, 2 \ldots K-1 \\ k \in K^{infeas} \end{array}$$

$$\left.\begin{aligned} \mu_B \geq \quad & L(x^{B_l}, y, \lambda^K, \mu^K)\big|_{x^k}^{lin} \\ & \nabla_{x_i} L(x, y, \lambda^K, \mu^K)\big|_{x^K} \leq 0 \quad if\ x_i^{B_l} = x_i^U \\ & \nabla_{x_i} L(x, y, \lambda^K, \mu^K)\big|_{x^K} \geq 0 \quad if\ x_i^{B_l} = x_i^L \end{aligned}\right\} \quad if\ K \in K^{feas}$$

$$\left.\begin{aligned} 0 \geq \quad & L_1(x^{B_l}, y, \lambda_1^K, \mu_1^K)\big|_{x^k}^{lin} \\ & \nabla_{x_i} L_1(x, y, \lambda_1^K, \mu_1^K)\big|_{x^K} \leq 0 \quad if\ x_i^{B_l} = x_i^U \\ & \nabla_{x_i} L_1(x, y, \lambda_1^K, \mu_1^K)\big|_{x^K} \geq 0 \quad if\ x_i^{B_l} = x_i^L \end{aligned}\right\} \quad if\ K \in K^{infeas}$$

Store the solution, if feasible, in $\mu_B^{stor}(K, B_l)$ and $y^{stor}(K, B_l)$.

(c) Select a new combination of bounds for x, say $B_l = B_2$.

(d) Repeat (b) and (c) until the relaxed dual problem has been solved at each set of bounds of the *connected* variables in x, that is, for every $B_l \in CB$.

STEP 4- *Selecting a new lower bound and* y^{K+1}*:*

From the stored set μ_B^{stor}, select the minimum μ_B^{min} (including the solutions from the current iteration). Also, select the corresponding stored value of $y^{stor}(k, B_j)$ as y^{min}. Set $M^{LBD} = \mu_B^{min}$, and $y^{K+1} = y^{min}$. Delete μ_B^{min} and y^{min} from the stored set.

STEP 5- *Check for convergence:*

Check if $M^{LBD} > P^{UBD} - \epsilon$. IF yes , STOP. Else, set $K = K + 1$ and return to step 1.

6.2 Automatic Implementation of the GOP algorithm

The **GOP** algorithm was implemented automatically through the use of the modeling language **GAMS** (Brooke *et al*, 1988). The actual solution of the convex primal and relaxed dual problems was achieved using **MINOS 5.2** (Murtagh and Saunders, 1988) on a **VAX**station 3200 and on a **MIPS RC2030** workstation, with **GAMS** as an interface.

In the automatic implementation of the **GOP** algorithm, one of the steps involves the selection of appropriate Lagrange functions from previous iterations to be present in the current *relaxed dual* problems. These are selected on the basis of satisfaction of the corresponding *qualifying* constraints. It is possible (especially in the case of iterations for which the primal problems are infeasible) that at some iteration, say the Kth one, the *qualifying* constraint of a Lagrange function from a previous iteration (say for the kth iteration) with respect to a particular x_j may be strictly satisfied as an equality. That is,

$$\nabla_{x_j} L(x, y = y^K, \lambda^k, \mu^k)\big|_{x^k} = 0$$

implying that all the Lagrange functions from the kth iteration with x_j set to either its lower or upper bound can be potentially present in the current *relaxed dual* problems. Thus, even if the *qualifying* constraints w.r.t the other x variables are satisfied as inequalities (that is, the *qualifying* constraints evaluated at y^K are either strictly less than or greater than zero), there will be two Lagrange functions from the kth iteration present in the *relaxed dual* problems of the current (Kth) iteration, along with their *qualifying* constraints. The *qualifying* constraints for these Lagrange functions w.r.t x_j are

$$\begin{aligned} \nabla_{x_j} L(x, y, \lambda^k, \mu^k)\big|_{x^k} &\geq 0 \quad if \;\; x_j^B = x_j^L \\ \nabla_{x_j} L(x, y, \lambda^k, \mu^k)\big|_{x^k} &\leq 0 \quad if \;\; x_j^B = x_j^U \end{aligned}$$

Together, these constraints imply that

$$\nabla_{x_j} L(x, y, \lambda^k, \mu^k) = 0$$

This reduces the degrees of freedom for the current *relaxed dual* problems by one, and could potentially lead to some regions of y being unavailable for the *relaxed dual* problems. This can be avoided if the *qualifying* constraints are introduced in a perturbed form as shown below :

$$\begin{aligned} \nabla_{x_j} L(x, y, \lambda^k, \mu^k) &\geq \delta \quad if \;\; x_j^B = x_j^L \\ \nabla_{x_j} L(x, y, \lambda^k, \mu^k) &\leq -\delta \quad if \;\; x_j^B = x_j^U \end{aligned}$$

where δ is a very small positive number. This ensures that when these *qualifying* constraints are checked at a later iteration, they cannot both be simultaneously satisfied. Hence, only one of the two corresponding Lagrange functions can be selected. The parameter δ can be made sufficiently small so that it does not significantly affect the optimal solution found by the *relaxed dual* problems.

6.3 GOP versus Global Optimum Search (GOS)

The Global Optimum Search (GOS) technique proposed by Floudas *et al* (1989) is based on the use of the Generalized Benders Decomposition algorithm (in conjunction with partitioning and transformations to remove nonconvexities) to solve nonconvex problems through a series of convex primal and master problems. The primal problem is solved in the same manner as in the **GOP** algorithm, that is, through projection on a subset of the variables and solution of the resulting convex subproblem.

It is in the master problem that the (GOS) differs from the **GOP** algorithm. While the **GOP** algorithm solves the relaxed dual problems for several regions of the y variables (by solving for different combinations of bounds for x), in the Global Optimum search, the x variables in the Lagrange function are fixed at the solution of the current primal problem. This is based on the assumption that

$$\min_{x \in X} L(x, y, \lambda^k, \mu^k) \geq L(x^k, y, \lambda^k, \mu^k)$$

where x^k is the solution of the kth primal problem for a fixed value of $y = y^k$. Since there is only one solution to the primal problem at every iteration, the (GOS) solves only one master problem at every iteration, in contrast to the **GOP** algorithm.

It should be noted, however, that fixing the x variables at the values corresponding to the solution in the primal problem is equivalent to linearization of the Lagrange function in the projected space of the y variables. Thus, there is an implicit assumption in the (GOS) technique that the optimal solutions of the primal problems for different values of $y = y^k$ are convex in y, thereby ensuring that the linearization of the Lagrange function underestimates the optimal solution to the problem for all values of y. In general, however, this cannot be guaranteed. This is the reason that the (GOS) technique, while often successful, does not have a theoretical guarantee of convergence to the global solution from any starting point.

7 Finite ϵ-Convergence and ϵ-Global Optimality

This section presents the theoretical proof of finite ϵ-convergence and ϵ-global optimality of the **GOP** algorithm.

Theorem 1: (Finite ϵ-convergence)

If the following conditions hold:

(a) *X and Y are nonempty compact convex sets,*

(b) *Conditions (A),*

(c) *$f(x, y)$, $g(x, y)$ and $h(x, y)$ are continuous on $X \times Y$, and*

(d) *The set $U(y)$ of optimal multipliers for the primal problem is nonempty for all $y \in Y$ and uniformly bounded in some neighbourhood of every such point,* then,

For any given $\epsilon > 0$, the **GOP** algorithm terminates in a finite number of steps .

Proof: Fix $\epsilon > 0$ arbitrarily. Suppose that the **GOP** algorithm does not converge in a finite number of iterations. Let $< y^k, \mu_B^{min^k} >$ be the sequence of optimal solutions to the relaxed dual problem at successive iterations k. By taking a subsequence, if necessary, we may assume that $< y^k, \mu_B^{min^k} >$ converges to $(\overline{y}, \overline{\mu_B^{min}})$ such that $\overline{y} \in Y$. At every iteration, there is an accumulation of constraints from previous iterations. Hence, the solution of any of the relaxed dual problems at a given iteration is greater than the μ_B^{min} for the previous iteration. This implies that μ_B^{min} is a nondecreasing sequence which is bounded above by the optimal value of the original problem. Also, at every iteration, y^k is in the compact set Y. Similarly, since $U(y)$ is uniformly bounded for all $y \in Y$, we may assume that the corresponding sequence of multipliers for the primal problems (λ^k, μ^k) converges to $(\overline{\lambda}, \overline{\mu})$. Let the solutions of the corresponding primal problems converge to $(\overline{x}, P(\overline{y})$). By Lemma 1 (see Appendix B) ,

$$L(x^{B_j}, \overline{y}, \overline{\lambda}, \overline{\mu})\Big|_{\overline{x}}^{lin} = L(\overline{x}, \overline{y}, \overline{\lambda}, \overline{\mu}) \quad for\ any\ x^{B_j} \tag{23}$$

Now, at every iteration k, due to accumulation of constraints,

$$\mu_B^k \geq L(x^{B_j}, y^{k+1}, \lambda^k, \mu^k)\Big|_{x^k}^{lin}$$

for some combination of bounds x^{B_j}. Therefore, by continuity of $L(x, y, \lambda^k, \mu^k)\Big|_{x^k}^{lin}$ and (23), we obtain $\overline{\mu_B^{min}} \geq L(\overline{x}, \overline{y}, \overline{\lambda}, \overline{\mu})$. Then, it only remains to be shown that $(\overline{\lambda}, \overline{\mu}) \in U(\overline{y})$, (Lemma 2 in Appendix B) for then, by the strong duality theorem, $L(\overline{x}, \overline{y}, \overline{\lambda}, \overline{\mu}) = P(\overline{y})$ and consequently, $\overline{\mu_B^{min}} \geq P(\overline{y})$; by the lower semicontinuity of $P(y)$ at $\overline{y}$ (Lemma 3 in Appendix B), this would imply that $\mu_B^{min^k} \geq P(y^k) - \epsilon$ for all k sufficiently large, which contradicts the assumption that the termination criterion in Step 5 is never met. □

Theorem 2: *(Global Optimality)*

If the following conditions hold:

(a) *X and Y are nonempty compact convex sets,*

(b) *Conditions (A) are satisfied, and*

(c) *Finite ϵ-convergence (Theorem 1),*

then,

(i) The solution of the Relaxed Dual **(RD)** problem in Step (3) will always be a valid underestimator of the solution of the problem (1).

(ii) The **GOP** algorithm will terminate at the global optimum of (1).

Proof:

(i) From Property 7, the solution of the relaxed dual problem in Step(3) (which is simply the solution of the RHS of (22)) will rigorously underestimate the solution of the relaxed dual problem (5). Since the relaxed dual problem (5) has fewer constraints than the original problem, it represents a lower bound on the solution of (1). Hence, the solution of the relaxed dual problem in Step (3) will always be a valid underestimator of the optimal solution of (1).

(ii) The primal problem at every iteration represents an upper bound for the original problem (1), while the relaxed dual problem contains fewer constraints than the original problem and thus represents a valid lower bound on the solution of (1). Therefore, since the termination of the algorithm is based on the difference between the lowest upper bound (from the primal problems) and the largest lower bound (from the relaxed dual problems), the algorithm will terminate when these two bounds are both within ϵ of the solution of (1). From Theorem 1, the algorithm terminates in a finite number of steps. Hence, the **GOP** algorithm terminates at an ϵ-global optimum of (1). □

8 Application to the Illustrating Example

This section considers the application of the **GOP** algorithm to the example problem that was used to illustrate the mathematical properties in section 5. This problem is given as

$$\min_{x,y} \; -x-y$$

$$\begin{aligned} s.t. \quad xy &\leq 4 \\ 0 \leq x &\leq 4 \\ 0 \leq y &\leq 8 \end{aligned}$$

This problem has a global solution of -8.5 at (0.5,8), and a local solution of -5 at (4,1). When the nonlinear programming solver **MINOS 5.2** (Murtagh and Saunders, 1988) was used to solve this seemingly simple problem, the global solution of -8.5 was found from some starting points. However, from a large number of starting points, **MINOS 5.2** failed to determine even a feasible solution.

Consider the starting point of $y = 2$ for the application of the **(GOP)** algorithm to the illustrating example.

Iteration 1 of the GOP algorithm

For a fixed value of $y^1 = 2$, the primal problem at the first iteration can be written as

$$\min_{x} -x-2$$

$$\begin{aligned} \textit{subject to} \quad 2x-4 &\leq 0 \\ 0-x &\leq 0 \\ x-4 &\leq 0 \end{aligned}$$

The solution of this problem yields $x = 2$, $\mu_1^1 = 0.5$, $\mu_2^1 = 0$, and $\mu_3^1 = 0$, where $\mu_1^1, \mu_2^1, \mu_3^1$ are the Lagrange multipliers corresponding to the above three constraints respectively for the first primal problem. The objective function has a value of -4, and provides an upper bound on the global solution. The Lagrange function formulated from this problem is

$$\begin{aligned} L(x, y, \mu^1) &= -x-y+\mu_1^1(xy-4)+\mu_2^1(0-x)+\mu_3^1(x-4) \\ &= -x-y+0.5(xy-4) \\ &= 0.5(y-2)x-y-2 \end{aligned}$$

Thus, the gradient of the Lagrange function w.r.t x is

$$\nabla_x L(x, y, \mu^1) = 0.5(y-2)$$

This is the form of the *qualifying* constraint to be added along with the Lagrange function to the *relaxed dual* problem.

At the first iteration, the two *relaxed dual* problems to be solved are shown below:

(i) For $y-2 \geq 0$, and $x^B = x^L = 0$:

$$\min_{y,\mu_B} \ \mu_B$$

$$\begin{aligned} \textit{subject to} \quad \mu_B &\geq L(x=0,y,\mu^1) = -y-2 \\ y-2 &\geq 0 \\ 0 \leq y &\leq 8 \end{aligned}$$

The solution of this problem is $y = 8, \ \mu_B = -10$.

(ii) For $y-2 \leq 0$, and $x^B = x^U = 4$:

$$\min_{y,\mu_B} \ \mu_B$$

$$\begin{aligned} \textit{subject to} \quad \mu_B &\geq L(x=4,y,\mu^1) = y-6 \\ y-2 &\leq 0 \\ 0 \leq y &\leq 8 \end{aligned}$$

The solution of this problem is $y = 0, \ \mu_B = -6$.

Thus, after the first iteration, there are two solutions of (μ_B, y) in the stored set. From these, the solution corresponding to the minimum μ_B is chosen. In this case, this corresponds to the solution $\mu_B = -10, \ y = 8$. Hence, the fixed value of y for the second iteration is 8 .

Illustration of GOP vs (GOS) :

It is interesting to note the consequence of using the (GOS) technique (Floudas *et al*, 1989) to solve this problem. The first primal problem is solved as in the **GOP** algorithm. However, in the Lagrange function formulated from this problem, the x variable is fixed at its solution in the primal problem, which is $x = 2$. This leads to

$$\begin{aligned} L(x=2,y,\mu^1) &= 0.5(y-2)2 - y - 2 \\ &= -4 \end{aligned}$$

From, this, it can be seen that the value of μ_B for the (GOS) master problem has to be greater than -4. This means that the (GOS) technique will converge after only one iteration at the suboptimal solution of $x = 2, \ y = 2$.

Iteration 2 of the GOP algorithm

For the second iteration, the primal problem is :

$$\min_x -x-8$$

$$\begin{aligned} \textit{subject to} \quad 8x-4 &\leq 0 \\ 0-x &\leq 0 \\ x-4 &\leq 0 \end{aligned}$$

This problem has a solution of $x = 0.5$, $\mu_1^2 = 0.125$, $\mu_2^2 = 0$, and $\mu_3^2 = 0$. The objective function value is -8.5, and thus provides a tighter upper bound than the solution from the first iteration.

The Lagrange function formulated from the second primal problem is

$$\begin{aligned} L(x, y, \mu^2) &= -x - y + 0.125(xy - 4) \\ &= (0.125y - 1)x - y - 0.5 \\ &= 0.125(y - 8)x - y - 0.5 \end{aligned}$$

Hence, the gradient of the Lagrange function *w.r.t.* x is

$$\nabla_x L(x, y, \mu^2) = 0.125(y - 8)$$

Thus, the *qualifying* constraint for this Lagrange function has the form

(i) For the problem with $x^B = x^U = 4$,

$$y \leq 8$$

(ii) For the problem with $x^B = x^L = 0$,

$$y \geq 8$$

Since the upper bound on y is 8, the second problem is nothing more than a function evaluation, and hence only the problem corresponding to $y \leq 8$ needs to be solved.

In order to determine the Lagrange function from the first iteration to be present in the current *relaxed dual* problem, the *qualifying* constraints for the Lagrange functions from the first iteration are checked for satisfaction at $y^2 = 8$. Since $y^1 = 2$, the constraint $y \geq y^1$ will be satisfied at $y = 8$. Hence, the first Lagrange function from the first iteration (corresponding to $x = 0$) is selected to be present in the current relaxed dual problem.

The relaxed dual problem for the second iteration is solved by setting $x = 4$:

$$\min_{y, \mu_B} \ \mu_B$$

$$\begin{aligned} \textit{subject to} \quad \mu_B &\geq L(x = 0, y, \mu^1) = -y - 2 \\ y - 2 &\geq 0 \\ \mu_B &\geq L(x = 4, y, \mu^2) = -0.5y - 4.5 \\ y - 8 &\leq 0 \\ 0 \leq y &\leq 8 \end{aligned}$$

The solution to this problem is $\mu_B = -8.5$, $y = 8$. Since this is a lower bound on the global solution, and is equal to the upper bound of -8.5 from the two primal problems, the algorithm stops, having determined the global solution after two iterations.

The algorithm was similarly applied to this problem from several starting points for y, and converged to the global solution in two iterations in each case.

9 Geometrical Interpretation

The application of the **GOP** algorithm to this example can be interpreted geometrically. For the starting point of $y_1^1 = 2$, the sequence of points generated by the algorithm is graphically illustrated in Figures 2(a)-2(d). Figure 2(a) shows the nonconvex feasible region, and Figure 2(b) represents $f(y)$, the optimal value of the primal problem for different fixed values of y.

For the first iteration (Figure 2(c)), with an optimal value of -4 for the primal problem, L_1^1 and L_2^1 are the Lagrangians evaluated at the two bounds $x = 0$ and $x = 4$ respectively. These Lagrange functions are underestimators of the objective function for $y \geq 2$ and $y \leq 2$ respectively, and they intersect at $y = 2$. It should be noted that these Lagrange functions intersect at $y = 2$ since the strong duality theorem is satisfied. The solutions of the two *relaxed dual* problems are at $y = 8, \mu_B = -10$ and $y = 0, \mu_B = -6$. The point $y = 8$ is chosen for the next iteration because it provides the lowest bound on the global solution.

At the second iteration, the fixed value of y is 8, and the primal problem has an objective value of -8.5 . One of the two *relaxed dual* problems to be solved in this iteration will be for $y \leq 8$. For this problem, there is a choice of two Lagrange functions L_1^1 and L_2^1 from the first iteration. Since the *qualifying* constraint for L_1^1 is $y \geq 2$, this is satisfied at $y = 8$. It can be seen from Figure 2(c) that L_1^1 underestimates the optimal value of the original problem for all values of y between $y^1 = 2$ and $y^2 = 8$, and therefore can be present in the *relaxed dual* problem for $y \leq 8$. In the case of L_2^1, however, it can be seen that this Lagrange function underestimates the optimal value of the original problem only for $y \leq 2$, and hence cannot be present for this *relaxed dual* problem.

The two relaxed dual problems solved at the second iteration each correspond to the region of y less than or greater than 8. The solution of the relaxed dual problem for $y \geq 8$ yields the solution $y = 8$, $\mu_B = -8.5$, since the upper bound for y is 8. The solution of the second relaxed dual problem, solved for $y \leq 8$, is also $y = 8$, $\mu_B = -8.5$. In this case, however, this happens because both the Lagrange function from the first iteration (L_1^1) and the Lagrange function from the second iteration for this problem (L_1^2) have the minimum *w.r.t* y at $y = 8$ (see Figure 2(d)). Hence, the algorithm stops after two iterations, having identified, and converged to, the global solution.

By storing the solutions of each of the relaxed dual problems at the current iteration, it is ensured that the algorithm can, if necessary, return a value of y from either side of y^K. The criterion for selecting the Lagrange functions from previous iterations results in the creation of an underestimating function that resembles a series of valleys and peaks, with the valleys representing the stored solutions of the relaxed dual problems at different iterations.

Finally, if the (GOS) technique were to be used to solve this problem from the same starting point, then the Lagrange function added to the master problem for the first iteration would be $\mu_B \geq -4$. From Figure 2(b), this is seen to be the linearization of the optimal solution of the problem around $y = 2$. This Lagrange function does not underestimate the optimal solution in *any* region of y.

10 Conclusions

In this paper, a new theoretical approach is proposed for the determination of global optima for several classes of nonconvex nonlinear programming **NLP** problems. These classes include problems with (a) quadratic objective function and linear constraints; (b) quadratic objective function and quadratic constraints; (c) polynomial functions; and (d) polynomial constraints. Theoretical properties have been investigated for the rigorous solution of the relaxed dual problem, and finite ϵ-convergence and ϵ-global optimality is proved. A global optimization algorithm, **GOP**, has been developed based upon these properties.

The scope for application of the **GOP** algorithm is not limited to just the classes of problems considered in this paper. It is possible to envisage the application of such an approach for problems involving integer quadratic programming, bilevel programming or linear/nonlinear complementarity problems. It can also be possibly used in conjunction with other approaches for solving problems where the structure can be exploited through decomposition. For example, most algorithms for solving mixed-integer nonlinear programming **MINLP** problems proceed through successive solution of **NLP** and **MILP** subproblems. In this context, the **GOP** algorithm could be used to determine a global solution of the **NLP** subproblem while efficient solvers are available for solving the **MILP** subproblem. Research work along these directions is currently in progress, and will be reported in future publications.

Acknowledgement

The authors gratefully acknowledge financial support from the National Science Foundation under Grant CBT-8857013, as well as support from Amoco Chemical Co., Tennessee Eastman Co., and Shell Development Co. .

Appendix A

Proof of Property 5: The linearization of the Lagrange function in the form given by (19) around the point x^k is

$$\begin{aligned} L(x, y, \lambda^k, \mu^k)\big|_{x^k}^{lin} &= L(x^k, y, \lambda^k, \mu^k) \\ &+ \left(\Phi_1(y, \lambda^k, \mu^k).\nabla_x \Psi_1(x)\big|_{x^k} + \nabla_x \Psi_2(x, \lambda^k, \mu^k)\big|_{x^k} \right).(x - x^k) \end{aligned}$$

From the **KKT** gradient conditions of the primal subproblem (2),

$$\nabla_x L(x^k, y^k, \lambda^k, \mu^k) = \Phi_1(y^k, \lambda^k, \mu^k).\nabla_x \Psi_1(x)\big|_{x^k} + \nabla_x \Psi_2(x, \lambda^k, \mu^k)\big|_{x^k} = 0.$$

Hence,

$$L(x, y, \lambda^k, \mu^k)\big|_{x^k}^{lin} = L(x^k, y, \lambda^k, \mu^k) + [\Phi_1(y, \lambda^k, \mu^k) - \Phi_1(y^k, \lambda^k, \mu^k)]\nabla_x \Psi_1(x)\big|_{x^k}(x - x^k) \tag{24}$$

Then, from (24) and Property 4, the Inner Relaxed Dual **(IRD)** Problem can be written as

$$\begin{aligned} L^*(\overline{x}, y, \lambda^k, \mu^k) &= \min_x L(x, y, \lambda^k, \mu^k) \\ &\geq \min_{x \in X} \{ L(x^k, y, \lambda^k, \mu^k) \ + \\ &\qquad [\, \Phi_1(y, \lambda^k, \mu^k) - \Phi_1(y^k, \lambda^k, \mu^k) \,] .\nabla_x \Psi_1(x)\big|_{x^k}.(x - x^k) \,\} \\ &\geq L(x^k, y, \lambda^k, \mu^k) \ + \\ &\qquad \min_{x \in X} \ \{ \, [\, \Phi_1(y, \lambda^k, \mu^k) - \Phi_1(y^k, \lambda^k, \mu^k) \,] .\nabla_x \Psi_1(x)\big|_{x^k}.(x - x^k) \,\} \end{aligned}$$

The last term on the right hand side is linearly separable in x_i. Therefore, this term can be written as

$$\begin{aligned} &\min_{x \in X} \quad \{ [\, \Phi_1(y, \lambda^k, \mu^k) - \ \Phi_1(y^k, \lambda^k, \mu^k) \,].\nabla_x \Psi_1(x^k)\big|_{x^k}.(x - x^k) \} \\ &= \min_{x \in X} \left\{ \sum_{i=1}^{n_1} \{ [\, \Phi_{1_i}(y, \lambda^k, \mu^k) - \Phi_{1_i}(y^k, \lambda^k, \mu^k) \,].\nabla_{x_i} \Psi_1(x^k)\big|_{x^k}.(x_i - x_i^k) \} \right\} \\ &= \sum_{i=1}^{n_1} \left\{ \min_{x_i} \{ [\, \Phi_{1_i}(y, \lambda^k, \mu^k) - \Phi_{1_i}(y^k, \lambda^k, \mu^k) \,].\nabla_{x_i} \Psi_1(x^k)\big|_{x^k}.(x_i - x_i^k) \} \right\} \end{aligned}$$

From this, it can be seen that if there exists some x_i such that $\Psi_1(x)$ is *not* a function of x_i, then the gradient of $\Psi_1(x)$ w.r.t that x_i will be zero. Therefore, the terms in the summation corresponding to this x_i will simply vanish, and the minimization of the Lagrange function w.r.t. these variables will not have any effect on the solution of the inner Relaxed Dual problem. Hence, the inner minimization problem, with the Lagrange function replaced by its linearization about x^k, needs to be considered only for those x_i that are in the function $\Psi_1(x)$. □

Proof of Property 6: Suppose that Property 6 is not true. Then, there should exist a value of the complicating variable $y^* \in Y$, such that

$$L(x, y^*, \lambda^k, \mu^k) < \min_{B_j \in CB} L(x^{B_j}, y^*, \lambda^k, \mu^k) \tag{25}$$

For $y = y^*$, $L(x, y^*, \lambda^k, \mu^k)$ is *linear* in x, which implies that the minimization of $L(x, y^*, \lambda^k, \mu^k)$ over x will result in a solution that lie at a bound of the compact set X. Thus, (25) cannot hold, and the hypothesis is not valid. Therefore, Property 6 must be true. □

Proof of Property 7 : From Property 4,

$$\min_{x \in X} L(x, y, \lambda^k, \mu^k) \geq \min_{x \in X} L(x, y, \lambda^k, \mu^k)\big|_{x^k}^{lin} \quad \forall y. \tag{26}$$

By its definition, then, $\overline{x}$ must satisfy the following inequality:

$$L^*(\overline{x}, y, \lambda^k, \mu^k) \geq \min_{x \in X} L(x, y, \lambda^k, \mu^k)\big|_{x^k}^{lin} \quad \forall y. \tag{27}$$

Using the definition of $L(x, y, \lambda^k, \mu^k)\big|_{x^k}^{lin}$, the right hand side of (27) is given by

$$\begin{aligned}
\min_{x \in X} L(x, y, \lambda^k, \mu^k)\big|_{x^k}^{lin} &= \min_{x \in X} \Big[L(x^k, y, \lambda^k, \mu^k) + \sum_{i=1}^{n} \nabla_{x_i} L(x, y, \lambda^k, \mu^k)\big|_{x^k} . (x_i - x_i^k)\Big] \\
&= L(x^k, y, \lambda^k, \mu^k) + \min_{x \in X} \sum_{i=1}^{n} \nabla_{x_i} L(x, y, \lambda^k, \mu^k)\big|_{x^k} . (x_i - x_i^k). \\
&\qquad\qquad \forall i \in I_c
\end{aligned}$$

Since the linearization of the Lagrange function is separable in x_i for any fixed $y = y^d$, the operators for minimization and summation can be exchanged. Hence,

$$\min_{x \in X} L(x, y, \lambda^k, \mu^k)\big|_{x^k}^{lin} = L(x^k, y, \lambda^k, \mu^k) + \sum_{i=1}^{n} \min_{x_i} \nabla_{x_i} L(x, y, \lambda^k, \mu^k)\big|_{x^k} . (x_i - x_i^k). \tag{28}$$

Consider the i th component of the second term on the right hand side. It is linear in x_i. Hence, the minimum of this term will lie at a bound of x_i, the specific nature of the bound(lower or upper) being determined by the sign of $\nabla_{x_i} L(x, y, \lambda^k, \mu^k)\big|_{x^k}$. Two cases are possible:

(a) If $\nabla_{x_i} L(x, y, \lambda^k, \mu^k)\big|_{x^k} \geq 0$, then

$$\min_{x_i} \nabla_{x_i} L(x, y, \lambda^k, \mu^k)\big|_{x^k} . (x_i - x_i^k) \geq \nabla_{x_i} L(x, y, \lambda^k, \mu^k)\big|_{x^k} . (x_i^L - x_i^k)$$

(b) If $\nabla_{x_i} L(x, y, \lambda^k, \mu^k)\big|_{x^k} \leq 0$, then

$$\min_{x_i} \nabla_{x_i} L(x, y, \lambda^k, \mu^k)\big|_{x^k} . (x_i - x_i^k) \geq \nabla_{x_i} L(x, y, \lambda^k, \mu^k)\big|_{x^k} . (x_i^U - x_i^k)$$

These two cases can be combined to yield the following result:

$$\begin{array}{l} \min_{x_i} \nabla_{x_i} L(x,y,\lambda^k,\mu^k)\big|_{x^k}.(x_i - x_i^k) \geq \nabla_{x_i} L(x,y,\lambda^k,\mu^k)\big|_{x^k}.(x_i^B - x_i^k) \\ where \\ \qquad x_i^B = \left\{ \begin{array}{ll} x_i^L & \forall y : \nabla_{x_i} L(x,y,\lambda^k,\mu^k)\big|_{x^k} \geq 0 \\ x_i^U & \forall y : \nabla_{x_i} L(x,y,\lambda^k,\mu^k)\big|_{x^k} \leq 0 \end{array} \right. \end{array} \tag{29}$$

Combining (27), (28) and (29), we see that

$$\begin{array}{rcl} L^*(\overline{x},y,\lambda^k,\mu^k) & \geq & L(x^k,y,\lambda^k,\mu^k) + \sum_{i=1}^{n_1} \nabla_{x_i} L(x,y,\lambda^k,\mu^k)\big|_{x^k}.(x_i^B - x_i^k) \\ & where & \\ & & x_i^B = x_i^L \quad \forall y : \ \nabla_{x_i} L(x,y,\lambda^k,\mu^k)\big|_{x^k} \geq 0 \\ & & x_i^B = x_i^U \quad \forall y : \ \nabla_{x_i} L(x,y,\lambda^k,\mu^k)\big|_{x^k} \leq 0 \\ & & \forall i \in I_c \end{array} \tag{30}$$

From this, it is evident that for any fixed or discretized $y = y^d$, there exists a combination of bounds B_j for the *connected* x variables such that

$$\begin{array}{rcl} \min_{x \in X} L(x,y^d,\lambda^k,\mu^k) & \geq & L(x^k,y^d,\lambda^k,\mu^k) + \sum_{i=1}^{n_1} \nabla_{x_i} L(x,y^d,\lambda^k,\mu^k)\big|_{x^k}.(x_i^{B_j} - x_i^k) \\ & \geq & L(x^{B_j},y^d,\lambda^k,\mu^k)\big|_{x^k}^{lin} \end{array}$$

Hence, for every discretized $y = y^d$, by fixing the values of the x variables at a combination of bounds B_j in the linearized Lagrange function and taking the minimum over all possible combinations of bounds $B_j \in CB$, a lower bound on the value of $L^*(\overline{x},y^d,\lambda^k,\mu^k)$ is obtained. Since this is true for every $y = y^d$, (21) must hold for all y. $\square$

Proof of Property 8 : The gradient of the Lagrange function (in the form given by (19)) is

$$\nabla_x L(x,y,\lambda^k,\mu^k)\big|_{x^k} = \Phi_1(y,\lambda^k,\mu^k).\nabla_x \Psi_1(x)\big|_{x^k} + \nabla_x \Psi_2(x,\lambda^k,\mu^k)\big|_{x^k}$$

By using the KKT conditions, this reduces to

$$\nabla_x L(x,y,\lambda^k,\mu^k)\big|_{x^k} = \left[\Phi_1(y,\lambda^k,\mu^k) - \Phi_1(y^k,\lambda^k,\mu^k)\right].\nabla_x \Psi_1(x)\big|_{x^k}$$

By assumption, $\Phi_1(y,\lambda^k,\mu^k)$ is linear in y. Therefore, since $\Phi_1(y^k,\lambda^k,\mu^k)$ and $\nabla_x \Psi_1(x)\big|_{x^k}$ are constant terms, $\nabla_x L(x,y,\lambda^k,\mu^k)\big|_{x^k}$ is linear in y. $\square$

Proof of Property 9 :

(a) For iteration 1:

For $k = 1$, from Property 7,

$$\min_{x \in X} L(x,y,\lambda^1,\mu^1) \geq \min_{B_l \in CB} \left\{ \begin{array}{c} L(x^{B_l},y,\lambda^1,\mu^1) \\ \nabla_{x_i} L(x,y,\lambda^1,\mu^1)\big|_{x^1} \leq 0 \quad if \ x_i^{B_l} = x_i^U \\ \nabla_{x_i} L(x,y,\lambda^1,\mu^1)\big|_{x^1} \geq 0 \quad if \ x_i^{B_l} = x_i^L \end{array} \right\}, \forall y.$$

Since this holds for all y,

$$\min_{y\in Y}\left\{\min_{x\in X} L(x,y,\lambda^1,\mu^1)\right\} \geq \min_{y\in Y}\left\{\min_{B_l\in CB}\left\{\begin{array}{c} L(x^{B_l},y,\lambda^1,\mu^1) \\ \nabla_{x_i}L(x,y,\lambda^1,\mu^1)\big|_{x^1}\leq 0 \quad if\ x_i^{B_l}=x_i^U \\ \nabla_{x_i}L(x,y,\lambda^1,\mu^1)\big|_{x^1}\geq 0 \quad if\ x_i^{B_l}=x_i^L \end{array}\right\}\right\}$$

The operators on the right hand side can be interchanged since $L(x^{B_l},y,\lambda^1,\mu^1)$ depends only on y. Therefore,

$$\min_{y\in Y}\left\{\min_{x\in X} L(x,y,\lambda^1,\mu^1)\right\} \geq \min_{B_l\in CB}\left\{\min_{y\in Y}\left\{\begin{array}{c} L(x^{B_l},y,\lambda^1,\mu^1) \\ \nabla_{x_i}L(x,y,\lambda^1,\mu^1)\big|_{x^1}\leq 0 \quad if\ x_i^{B_l}=x_i^U \\ \nabla_{x_i}L(x,y,\lambda^1,\mu^1)\big|_{x^1}\geq 0 \quad if\ x_i^{B_l}=x_i^L \end{array}\right\}\right\}$$

Equivalently, this can be written as

$$\left\{\begin{array}{l} \min\limits_{\substack{y\in Y\\ \mu_B}} \ \mu_B \\ s.t. \quad \mu_B \geq \min\limits_{x\in X} L(x,y,\lambda^1,\mu^1) \end{array}\right\} \geq \min_{B_l\in CB}\left\{\begin{array}{l} \min\limits_{y\in Y,\mu_B} \ \mu_B \\ s.t. \quad \mu_B \geq L(x^{B_l},y,\lambda^1,\mu^1) \\ \nabla_{x_i}L(x,y,\lambda^1,\mu^1)\big|_{x^1}\leq 0 \quad if\ x_i^{B_l}=x_i^U \\ \nabla_{x_i}L(x,y,\lambda^1,\mu^1)\big|_{x^1}\geq 0 \quad if\ x_i^{B_l}=x_i^L \end{array}\right\}$$

Hence, for iteration 1, the property is proved.

(b) For iteration 2:

$UL(1,2)$ represents the set of Lagrange functions from the first iteration whose qualifying constraints are satisfied at y^2. For the second iteration, the Lagrange functions from the first iteration that belong to $UL(1,2)$ already have the value of x set to the appropriate bound. Therefore, these Lagrangians are functions only of y.

Thus, we need to show that if

$$\overline{\mu}_B{}^2 = \left\{\begin{array}{l} \min\limits_{y,\mu_B} \ \mu_B \\ \\ s.t. \\ \quad \left.\begin{array}{ll} \mu_B \geq & L(x^{B_j},y,\lambda^1,\mu^1)\big|_{x^1}^{lin} \\ & \nabla_{x_i}L(x,y,\lambda^1,\mu^1)\big|_{x^1}\leq 0 \quad if\ x_i^{B_j}=x_i^U \\ & \nabla_{x_i}L(x,y,\lambda^1,\mu^1)\big|_{x^1}\geq 0 \quad if\ x_i^{B_j}=x_i^L \end{array}\right\} \forall j\in UL(1,2) \\ \\ \quad \mu_B \geq \min\limits_{x\in X} \ L(x,y,\lambda^2,\mu^2) \end{array}\right\} \tag{31}$$

Then,

$$\overline{\mu}_B{}^2 \geq \min_{B_l\in CB}\left\{\begin{array}{l} \min\limits_{y,\mu_B} \ \mu_B \\ \\ s.t. \\ \quad \left.\begin{array}{ll} \mu_B \geq & L(x^{B_j},y,\lambda^1,\mu^1)\big|_{x^1}^{lin} \\ & \nabla_{x_i}L(x,y,\lambda^1,\mu^1)\big|_{x^1}\leq 0 \quad if\ x_i^{B_j}=x_i^U \\ & \nabla_{x_i}L(x,y,\lambda^1,\mu^1)\big|_{x^1}\geq 0 \quad if\ x_i^{B_j}=x_i^L \end{array}\right\} \forall j\in UL(1,2) \\ \\ \quad \begin{array}{ll} \mu_B \geq & L(x^{B_l},y,\lambda^2,\mu^2)\big|_{x^2}^{lin} \\ & \nabla_{x_i}L(x,y,\lambda^2,\mu^2)\big|_{x^2}\leq 0 \quad if\ x_i^{B_l}=x_i^U \\ & \nabla_{x_i}L(x,y,\lambda^2,\mu^2)\big|_{x^2}\geq 0 \quad if\ x_i^{B_l}=x_i^L \end{array} \end{array}\right\} \tag{32}$$

Since the first set of constraints for this problem are functions of only y, the $\min\limits_{B_l \in CB}$ operator applies only to the second set of constraints, i.e., those corresponding to the 2^{nd} primal problem. Hence, (32) is equivalent to

$$\overline{\mu}_B{}^2 \geq \left\{ \begin{array}{l} \min\limits_{y,\mu_B} \;\; \mu_B \\ \\ s.t. \\ \quad \left. \begin{array}{rl} \mu_B \geq & L(x^{B_j}, y, \lambda^1, \mu^1)\big|_{x^1}^{lin} \\ & \nabla_{x_i} L(x, y, \lambda^1, \mu^1)\big|_{x^1} \leq 0 \quad if\ x_i^{B_j} = x_i^U \\ & \nabla_{x_i} L(x, y, \lambda^1, \mu^1)\big|_{x^1} \geq 0 \quad if\ x_i^{B_j} = x_i^L \end{array} \right\} \forall j \in UL(1,2) \\ \\ \quad \mu_B \geq \min\limits_{B_l \in CB} \left\{ \begin{array}{l} L(x^{B_l}, y, \lambda^2, \mu^2)\big|_{x^2}^{lin} \\ \nabla_{x_i} L(x, y, \lambda^2, \mu^2)\big|_{x^2} \leq 0 \quad if\ x_i^{B_l} = x_i^U \\ \nabla_{x_i} L(x, y, \lambda^2, \mu^2)\big|_{x^2} \geq 0 \quad if\ x_i^{B_l} = x_i^L \end{array} \right. \end{array} \right\} \tag{33}$$

The use of Property 7 for the second iteration gives the following result:

$$\min_{x \in X} L(x, y, \lambda^2, \mu^2) \;\geq\; \min_{B_l \in CB} \left\{ \begin{array}{l} L(x^{B_l}, y, \lambda^2, \mu^2) \\ \nabla_{x_i} L(x, y, \lambda^2, \mu^2)\big|_{x^2} \leq 0 \quad if\ x_i^{B_l} = x_i^U \\ \nabla_{x_i} L(x, y, \lambda^2, \mu^2)\big|_{x^2} \geq 0 \quad if\ x_i^{B_l} = x_i^L \end{array} \right\} \tag{34}$$

Hence, from (33) and (34), the second set of constraints on the RHS of (32) is simply a relaxed form of the second set of constraints of the RHS of (31). Hence, for $k = 2$, the property holds. Similarly, by induction, the property can be proved for any k. □

Appendix B

Lemma 1: If the sequence of solutions $< y^k >$ of the relaxed dual problem converges to $\overline{y}$, and the corresponding Lagrange multipliers for the primal problem **(P)** converge to $(\overline{\lambda}, \overline{\mu})$, then

$$L(x^{B_j}, \overline{y}, \overline{\lambda}, \overline{\mu})\big|_{\overline{x}}^{lin} = L(\overline{x}, \overline{y}, \overline{\lambda}, \overline{\mu}) \quad for\ any\ x^{B_j}$$

Proof: As $y^k \rightarrow \overline{y}$, we may assume that $x^k \rightarrow \overline{x}$, $\lambda^k \rightarrow \overline{\lambda}$ and $\mu^k \rightarrow \overline{\mu}$. Now, the linearization of the Lagrangian as given by (24) and evaluated at x^{B_j} is

$$\begin{array}{rcl} L(x^{B_j}, y, \lambda^k, \mu^k)\big|_{x^k}^{lin} & = & L(x^k, y, \lambda^k, \mu^k) \\ & & +[\Phi_1(y, \lambda^k, \mu^k) - \Phi_1(y^k, \lambda^k, \mu^k)].\nabla_x \Psi_1(x)\big|_{x^k}.(x^{B_j} - x^k) \\ & & \qquad\qquad \forall B_j \in CB \end{array}$$

The Lagrange function is a weighted sum of $f(x,y)$, $g(x,y)$ and $h(x,y)$, all of which are continuous over $X \times Y$. Therefore, $L(x, y, \lambda^k, \mu^k)$ is also continuous over $X \times Y$ and consequently, $L(x^k, y, \lambda^k, \mu^k)$ is continuous over Y. This implies that it is also *lower semicontinuous* at every y, and therefore at $\overline{y}$. Therefore,

$$As \quad y^k \rightarrow \overline{y}, \qquad L(x^k, y, \lambda^k, \mu^k)\big|_{y=\overline{y}} \rightarrow L(\overline{x}, \overline{y}, \overline{\lambda}, \overline{\mu}) \tag{35}$$

The function $\Phi_1(y, \lambda^k, \mu^k)$, is a linear function of y, and hence it is continuous over y. Therefore, it is *lower semicontinuous* at $\overline{y}$, which implies that

$$As \quad y^k \to \overline{y}, \qquad \Phi_1(y, \lambda^k, \mu^k)\big|_{y=y^k} \to \Phi_1(\overline{y}, \overline{\lambda}, \overline{\mu}) \tag{36}$$

But as $y^k \to \overline{y}$, $\Phi_1(y^k, \lambda^k, \mu^k) \to \Phi_1(\overline{y}, \overline{\lambda}, \overline{\mu})$. From this, and (35) and (36) we obtain the desired result. □

Lemma 2: For a fixed $y = y^k$ for the primal problem, let $U(y^k)$ be the set of optimal multipliers for the primal problem. If $y^k \to \overline{y}$ and $\lambda^k \to \overline{\lambda}, \mu^k \to \overline{\mu}$, then $(\overline{\lambda}, \overline{\mu}) \in U(\overline{y})$.
Proof: The proof for this Lemma comes from showing that $U(y)$ is an upper semicontinuous mapping at $\overline{y}$. To do this, we employ the characterization of $U(y)$ as the set of optimal solutions of the dual of the primal problem; that is ,

$$U(y) = \{\mu \geq 0, \lambda : \; L^*(\overline{x}, y, \lambda, \mu) = \max_{\mu^1 \geq 0, \lambda^1} L^*(\overline{x}, y, \lambda^1, \mu^1)\}$$

Now, $L^*(\overline{x}, y, \lambda, \mu)$ is a continuous function, since it is a sum of linear continuous functions of y. Then, application of Theorem 1.5 of Meyer (1970) proves the desired result. □

Lemma 3: Let $P(y)$ be the set of optimal solutions of the primal problem for fixed values of y; then, if $y^k \to \overline{y}$, $P(y)$ is lower semicontinuous at $\overline{y}$.
Proof: $P(y)$ is the solution of the primal problem for a fixed value of y; therefore, since the strong duality theorem can be applied for every primal problem, it is also equal to the optimal value of the corresponding dual problem. Since the set $U(y)$ is assumed to be nonempty for all y, this implies that

$$P(y) = \max_{\mu \geq 0, \lambda} L^*(\overline{x}, y, \lambda, \mu) \quad \forall y \in Y$$

Since Y is a compact set, the local uniform boundedness of $U(y)$ implies the uniform boundedness of $U(y)$ on all of Y. Therefore, there exists a compact set U^* such that $U(y) \subseteq U^*$ for all y in Y. The constraint $(\lambda, \mu) \in U^*$ can then be added without disturbing the equality, and this allows the direct application of Lemma 1.2 of Meyer (1970) in order to obtain the desired lower semicontinuity of $P(y)$. □

References

[1] Aggarwal, A., and Floudas, C.A., Global Optimum Search in Quadratic Programming, *Presented at the CORS/TIMS/ORSA Joint National Meeting, Vancouver, Canada* (1989).

[2] Aggarwal, A., and Floudas, C.A., A Decomposition Approach for Global Optimum Search in QP, NLP and MINLP problems, *Annals of Operations Research*, in Press, (1990).

[3] Al-Khayyal, F.A., and Falk, J.E., Jointly Constrained Biconvex Programming, *Mathematics of Operations Research*, **8**, (1983).

[4] Archetti, F., and Schoen, F., A Survey on the Global Minimization Problem : General Theory and Computational Approaches, *Annals of Operations Research*, **1**, 87 (1984).

[5] Avriel, M., *Nonlinear Programming : Analysis and Methods*, Prentice-Hall Inc., Engelwood Cliffs, New Jersey (1976).

[6] Branin, F.H., Widely Convergent Methods for Finding Multiple Solutions of Simultaneous Nonlinear Equations, *IBM Journal of Research Developments*, 504 (1972).

[7] Brooke, A., Kendrick, D., and Meeraus, M, *GAMS: A User's Guide.* , Scientific Press, Redwood City (1988).

[8] Dixon, L.C.W., and Szego, G.P., Editors, *Towards Global Optimization*, North Holland, Amsterdam (1975).

[9] Dixon, L.C.W., and Szego, G.P., Editors, *Towards Global Optimization 2.*, North-Holland, Amsterdam (1978).

[10] Floudas, C.A., and Aggarwal, A., A Decomposition Strategy for Global Optimum Search in the Pooling Problem, *Operations Research Journal on Computing*, **2(3)**, in Press (1990).

[11] Floudas, C.A., Aggarwal, A., and Ciric, A.R., Global Optimum Search for Nonconvex NLP and MINLP problems, *Computers and Chemical Engineering*, **13**, 1117 (1989).

[12] Floudas, C.A., and Pardalos, P.M., A Collection of Test Problems for Constrained Global Optimization Algorithms, *Lecture Notes in Computer Science*, Springer-Verlag, in Press (1990).

[13] Geoffrion, A.M., Generalized Benders Decomposition, *Journal of Optimization Theory and Applications*, **10**, 237 (1972).

[14] Hansen, E.R., Global Optimization using Interval Analysis: The One-Dimensional Case, *Journal of Optimization Theory and Applications*, **29**, 331 (1979).

[15] Hansen, P., Jaumard, B., and Lu, S-H., Global Optimization of Univariate Lipschitz functions : I. Survey and Properties, *RUTCOR Research Report 18-89*, Rutgers University (1989a).

[16] Hansen, P., Jaumard, B., and Lu, S-H., Global Optimization of Univariate Lipschitz functions : II. New Algorithms and Computational Comparisons, *RUTCOR Research Report 23-89*, Rutgers University (1989b).

[17] Hansen, P., Jaumard, B., and Lu, S-H., An Analytical Approach to Global Optimization, *To appear in Mathematical Programming* (1990).

[18] Horst, R., and Tuy, H., On the Convergence of Global Methods in Multiextremal Optimization, *Journal of Optimization Theory and Applications*, **54**, 253 (1987).

[19] Horst, R., and Tuy, H., *Global Optimization: Deterministic Approaches*, Springer-Verlag (1990).

[20] Kirkpatrick, S., Gelatt, C.D, and Vecchi, M.P., Optimization by Simulated Annealing, *Science*, **220**, 671 (1983).

[21] Kocis, G.R., and Grossmann, I.E., Global Optimization of Nonconvex **MINLP** Problems in Process Synthesis, *Industrial and Engineering Chemistry Research*, **27**, 1407 (1988).

[22] Kocis, G.R., and Grossmann, I.E., Computational Experience with **DICOPT** solving **MINLP** problems in Process Systems Engineering, *Computers and Chemical Engineering*, **13**, 307 (1989).

[23] Levy, A.V., and Montalvo, A., The Tunneling Algorithm for the Global Minimization of Functions , *SIAM J. of Sci. Stat. Comp.*, **6**, 15 (1985).

[24] Meyer, R., The Validity of a Family of Optimization Methods, *SIAM Journal on Control*, **8**, 41 (1970).

[25] Murtagh, B.A., and Saunders, M.A., *MINOS 5.0 User's Guide – Appendix A: MINOS 5.2.* , Technical Report SOL 83-20R, Systems Optimization Laboratory, Dept. of Operations Research, Stanford University (1988).

[26] Murty, K.G. and Kabadi, S.N., Some NP-Complete Problems in Quadratic and Nonlinear Programming, *Mathematical Programming*, **39**, 117 (1987).

[27] Pardalos, P.M., and Rosen, J.B., Methods for Global Concave Minimization: A Bibliographic Survey, *SIAM Review*, **28**, 367 (1986).

[28] Pardalos, P.M., and Rosen, J.B., Constrained Global Optimization : Algorithms and Applications, *v. 268 of Lecture Notes in Computer Science*, Springer Verlag (1987).

[29] Pardalos, P.M. and Schnitger, G., Checking Local Optima for Constrained Quadratic Programing is NP-hard, *Operations Research Letters*, **7**, 33 (1988).

[30] Pinter, J., Convergence Properties of Stochastic Optimization Procedures, *Math. Operat. Stat., Ser. Optimization*, **15**, 405 (1984).

[31] Piyavskii, S.A., An Algorithm for finding the Absolute Extremum of a Function, *USSR Comput. Math. Phys.*, **12**, 57 (1972).

[32] Ratschek, H., and Rokne, J., *New Computer Methods for Global Optimization*, Halsted Press (1988).

[33] Rinnoy Kan, A.H.G., and Timmer, G.T., Stochastic Global Optimization Methods. Part I: Clustering Methods, *Mathematical Programming*, **39**, 27 (1987).

[34] Stephanopoulos, G., and Westerberg, A.W., The Use of Hestenes' Method of Multipliers to Resolve Dual Gaps in Engineering System Optimization, *Journal of Optimization Theory and Applications*, **15**, 285 (1975).

[35] Torn, A., and Zilinskas, A., Global Optimization , *Lecture Notes in Computer Science, No. 350*, Springer-Verlag (1987).

[36] Tuy, H., Thieu, T.V., and Thai, N.Q., A Conical Algorithm for Globally Minimizing a Concave Function Over a Closed Convex Set, *Mathematics of Operations Research*, **10**, 498 (1985).

[37] Visweswaran, V., and Floudas, C.A., A Global Optimization Algorithm (**GOP**) for Certain Classes of Nonconvex NLPs : II. Application of Theory and Test Problems, *Submitted for Publication* (1990).

[38] Westerberg, A.W. and Shah, J.V., Assuring a Global Minimum by the Use of an Upper Bound on the Lower (Dual) Bound, *Computers and Chemical Engineering*, **2**, 83 (1978).

[39] Wingo, D.R., Globally Minimizing Polynomials Without Evaluating Derivatives, *International Journal of Computer Mathematics*, **17**, 287 (1985).

[40] Zilinskas, A., *Global Optimization – Axiomatics of Statistical Models, Algorithms and Their Applications*, Moklas, Vilnius (1986)