

1990

# Convergence properties of generalized Benders decompositions

N. V. Sahinidis  
*Carnegie Mellon University*

Ignacio E. Grossmann

Carnegie Mellon University. Engineering Design Research Center.

Follow this and additional works at: <http://repository.cmu.edu/cheme>

---

This Technical Report is brought to you for free and open access by the Carnegie Institute of Technology at Research Showcase @ CMU. It has been accepted for inclusion in Department of Chemical Engineering by an authorized administrator of Research Showcase @ CMU. For more information, please contact [research-showcase@andrew.cmu.edu](mailto:research-showcase@andrew.cmu.edu).

**NOTICE WARNING CONCERNING COPYRIGHT RESTRICTIONS:**  
The copyright law of the United States (title 17, U.S. Code) governs the making of photocopies or other reproductions of copyrighted material. Any copying of this document without permission of its author may be prohibited by law.

**Convergence Properties of  
Generalized Benders Decompositions**

by

N.V. Sahinidis, I. E. Grossmann

EDRC 06-89-90 C . 3

**CONVERGENCE PROPERTIES OF  
GENERALIZED BENDERS DECOMPOSITION**

N.V. Sahinidis and I.E. Grossmann\*

Department of Chemical Engineering  
Carnegie Mellon University  
Pittsburgh, PA 15213

May 1990

---

\*Author to whom correspondence should be addressed.

## **ABSTRACT**

This paper addresses two major issues related to the convergence of generalized Benders decomposition which is an algorithm for the solution of mixed integer linear and nonlinear programming problems. First, it is proved that a mixed integer nonlinear programming formulation with zero nonlinear programming relaxation gap requires only one Benders cut in order to converge, namely the cut corresponding to the optimal solution. This property indicates the importance of developing tight formulations for integer programs. Second, it is demonstrated that the application of generalized Benders decomposition to nonconvex problems does not always lead to the global optimum for these problems; it may not even lead to a local optimum. It is shown that this property follows from the fact that every local optimum of a nonlinear program gives rise to a local optimum in the projected problem of Benders. Examples are given to illustrate the properties.

## 1. Introduction

A large number of chemical engineering problems can be formulated as mathematical programming problems of the form:

$$\underset{x, y}{\text{minimize}} \quad f(x, y) \quad \text{subject to} \quad G(x, y) \leq 0 \quad x \in X \quad y \in Y \quad (1)$$

where  $f: X \times Y \rightarrow \mathbb{R}$ ,  $G: X \times Y \rightarrow \mathbb{R}^m$ , and  $X \times Y \subset \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$ . It is assumed that  $y$  is a vector of *complicating variables*: variables which, when temporarily fixed, render the remaining optimization problem in  $x$  considerably more tractable. This particularly applies to the following situations (Geoffrion, 1972):

(a) For fixed  $y$ , (1) separates into a number of independent optimization problems, each involving a different subvector of  $x$ . Typical examples of this case are multiperiod design problems.

(b) For fixed  $y$ , (1) assumes a well-known special structure for which efficient solution procedures are available. Examples of this case are both mixed integer linear and mixed integer nonlinear programming problems in which fixing the integer variables ( $y$ ) gives rise to a linear and nonlinear programming problem, respectively.

(c) Problem (1) is not a convex program in  $x$  and  $y$  jointly, but fixing  $y$  renders it so in  $x$ . Nonlinear programming problems involving bilinear terms of the form  $(xy)$  fall into this category, since by fixing the  $y$ -variables the nonlinearities are removed.

The central objective in all the above cases is to somehow exploit the special structure. By looking at the problem in the  $y$ -space rather than in the  $xy$ -space, we expect that, in situation (a), the computations can be done in parallel for each of the smaller independent subproblems; in (b), use can be made of efficient special purpose algorithms; and in (c), the nonconvexities can be treated separately from the convex portion of the problem.

Benders (1962) devised an approach for exploiting special structure in the class of problems where, fixing the complicating variables reduces the given problem to an ordinary linear program, parameterized of course by the value of the complicating vector. His approach builds adequate representations of (i) the extremal value of the linear program as a function of the parameterizing vector, and (ii) the set of values of the parameterizing vector for which the linear program is feasible. Linear programming duality theory was employed to derive these representations and led to the derivation of the *master problem*: a new problem equivalent to the original one. Then, a relaxation solution strategy was suggested to solve this problem. The relaxed master problem is solved to yield a lower bound for the minimization problem and to determine values for the y-variables. Subsequently, for fixed values of the complicating variables, the resulting linear programming subproblem yields an upper bound for the objective and dual information to construct *Benders cuts*: constraints which augment the relaxed representation of the master problem. The procedure terminates when the lower and upper bounds become sufficiently close.

Geoffrion (1972) generalized Benders<sup>1</sup> approach to a broader class of programs in which the parameterized subproblem need no longer be a linear program. Nonlinear convex duality theory was employed to derive the equivalent master problem. In this *generalized Benders decomposition* (GBD), the algorithm alternates between the solution of relaxed master problems and convex nonlinear subproblems.

Benders decomposition (generalized or not) has been applied to a variety of problems that were modelled as mixed integer linear programming (MDLP) or mixed integer nonlinear programming (MINLP) problems. Geoffrion and Graves (1974) were among the first to use the algorithm to solve an MILP model for the design industrial distribution systems. Noonan and Giglio (1977) used Benders decomposition coupled with a successive linearization procedure to solve a nonlinear multiperiod mixed integer model for planning the expansion of electric power generation networks. Rouhani *et al* (1985) used generalized Benders

decomposition to solve an MINLP model for reactive source planning in power systems. Floudas and Ciric (1989) applied GBD to their MINLP formulation for heat exchanger network synthesis. El-Halwagi and Manouthisakis (1989) developed an MINLP model for the design and analysis of mass exchange networks and suggested the use of GBD for its solution. Sahinidis and Grossmann (1990b) applied the algorithm to the problem of scheduling multiproduct parallel production lines by solving the corresponding large scale MINLP model. The technique has also been used for solving nonconvex nonlinear programming (NLP) and MINLP problems: Geoffrion (1972) applied it to the variable factor programming problem and Floudas *et al* (1989) suggested a Benders-based procedure for searching for the global optimum of nonconvex problems.

Unfortunately, Benders decomposition has not been uniformly successful in all its applications. Florian *et al* (1976) have noticed that the algorithm often converged very slowly when applied to their MILP model for scheduling the movement of railway engines. Bazaraa and Sherali (1980) observed that a large number of iterations were needed to solve their MILP model for quadratic assignment problems of realistic size. Sahinidis *et al* (1989) have found branch and bound to be significantly faster than Benders decomposition for the solution of their multiperiod MILP for long range planning in the chemical industries. Finally, with respect to the application of GBD to nonconvex problems, in contrast to Geoffrion (1972) and Floudas *et al* (1989), Sahinidis and Grossmann (1989) have encountered cases where the application of this technique to their MINLP model for the planning of chemical processes failed to produce the global optima of these problems.

This paper discusses two major issues related to the convergence of generalized Benders decomposition. In Section 2, we present a rigorous derivation of the algorithm and review known theoretical results regarding convergence properties and the conditions under which the technique is applicable. In Section 3, we prove that a mixed integer nonlinear programming formulation with zero nonlinear programming relaxation gap requires only one

Benders cut in order to converge. The result provides some insights on the impact of problem formulation on the behavior of the GBD method. Finally, in Section 4, it is shown that the application of generalized Benders decomposition to nonconvex problems does not always lead to the global optimum for these problems and that it may not even lead to a local optimum. The reasons why this happens are also discussed.

## 2. Derivation of Generalized Benders Decomposition

The key idea which enables (1) to be viewed as a problem in  $y$ -space is the concept of *projection* (Geoffrion, 1970). The projection of (1) onto  $y$  is:

$$\underset{y}{\text{minimize}} \quad v(y) \quad \text{subject to} \quad y \in Y \cap V \quad (2)$$

where

$$v(y) = \underset{x}{\text{infimum}} [ f(x, y) \quad \text{subject to} \quad G(x, y) \leq 0, \quad x \in X ] \quad (3)$$

and

$$V \equiv \{y: G(x, y) \leq 0 \quad \text{for some } x \in X\} \quad (4).$$

Note that  $v(y)$  is the optimal value of (1) for fixed  $y$  and that, by designating  $y$  as the complicating variables, evaluating  $v(y)$  is much easier than solving (1) itself. The set  $V$  consists of those values of  $y$  for which (3) is feasible;  $Y \cap V$  is the projection of the feasible region of (1) onto the  $y$ -space. The following theorem (Geoffrion, 1970) shows that the projected problem (2) is equivalent to the original problem (1):

**Theorem 1 (Projection):** Problem (1) is infeasible or has unbounded optimal value if and only if the same is true of (2). If  $(x^*, y^*)$  is optimal in (1), then  $y^*$  is optimal in (2). If  $y^*$  is optimal in (2) and  $x^*$  achieves the infimum in (3) with  $y = y^*$ , then  $(x^*, y^*)$  is optimal in (1). If  $y'$  is  $\epsilon_1$ -optimal in (2) and  $x'$  is  $\epsilon_2$ -optimal in (3) with  $y = y'$ , then  $(x', y')$  is  $(\epsilon_1 + \epsilon_2)$ -optimal in (1).

In light of this theorem and the assumption that the optimization problem in (3) is easy, it is important to consider (2) as a route of solving (1). However, the difficulty with (2) is that the function  $x^*$  and the set  $V$  are only known implicitly via their definitions (3) and (4). In order to resolve this issue, the following two manipulations are going to be applied to (2): (i) invoke the natural dual representation of  $V$  in terms of the intersection of a collection of regions which contain it; and (ii) invoke the natural dual representation of  $x^*$  in terms of the pointwise supremum of a collection of functions which dominate it. These manipulations are based on the two following theorems due to Geoffrion (1972):

**Theorem 2 (V-Representation):** Assume that  $X$  is a nonempty convex set and that  $G$  is convex on  $X$  for each fixed  $y \in Y$ . Assume further that the set  $Z(y) = \{z \in \mathbb{R}^m : G(x,y) \leq z \text{ for some } x \in X\}$  is closed for each fixed  $y$ . Then, a point  $y \in Y$  is also in the set  $V$  if and only if  $y$  satisfies the (infinite) system

$$\left[ \inf_{x \in X} k^T G(x,y) \right] \leq 0 \quad \text{all } y \in A \quad (5)$$

where:

$$A = \left\{ y \in Y : X(y) \neq \emptyset \text{ and } \inf_{x \in X} G(x,y) = 0 \right\} \quad (6).$$

**Theorem 3 ( $\wedge$ -Representation):** Assume that  $X$  is a nonempty convex set and that  $f$  and  $G$  are convex on  $X$  for each fixed  $y \in Y$ . Assume further that, for each fixed  $y \in Y \cap V$ , at least one of the following three conditions holds: (a)  $v(y^*)$  is finite and the optimization problem in (3) possesses an optimal multiplier vector, (b)  $v(y^*)$  is finite,  $G(x, y^*)$  and  $f(x, y^*)$  are continuous on  $X$ ,  $X$  is closed, and the  $\epsilon$ -optimal solution set of the optimization problem in (3) with  $y = y^*$  is nonempty and bounded for some  $\epsilon \geq 0$ ; and (c)  $v(y^*) = -\infty$ , i.e. (3) is unbounded. Then, the optimal value of the optimization problem in (3) equals that of its dual on  $Y \cap V$ , that is:

$$v(y) = \sup_{u \geq 0} [\inf_{x \in X} f(x, y) + u^T G(x, y)] \quad y \in Y \cap V \quad (7).$$

The proof of the last theorem is essentially based on strong duality theory (see Bazaraa and Shetty, 1979). Under this, and the rest of the assumptions of Theorems 2 and 3, the projection and dualization manipulations applied to (1) yield the following equivalent master **problem:**

$$\text{minimize } [ \sup_{y \in Y} \{ \inf_{u \geq 0} \{ \inf_{x \in X} [ f(x, y) + u^T G(x, y) ] \text{ s.t. (5) } \} } ] \quad (8)$$

which by using the definition of the supremum as the lowest upper bound Geoffrion (1972) recasts as follows (assuming convexity of  $f(x, y)$ ) as will become clear in Section 4 of the paper):

Problem (M):

$$\text{minimize } r \quad (9)$$

$$\eta \in \mathcal{M} \quad y \in Y$$

s.t.

$$T_i \geq \inf_{x \in X} [ f(x, y) + u^T G(x, y) ] \quad \text{all } u \geq 0 \quad (10)$$

$$[ \inf_{x \in X} \lambda^T G(x, y) ] \leq 0 \quad \text{all } X \in A \quad (5)$$

The above master problem is a semi-infinite optimization problem since it involves an infinite number of constraints. Therefore, *relaxation* is the most natural strategy for solving it. Begin by solving a relaxed version of Problem (M) which ignores all but a few of the constraints (10) and (5); if the resulting solution does not satisfy all of the ignored constraints, then generate and add to the relaxed problem one or more violated constraints and solve it again; continue in this fashion until a relaxed problem solution satisfies all of the ignored

constraints, or until a termination criterion signals that a solution of acceptable accuracy has been obtained.

Suppose that  $(y^*, r|^*)$  is optimal in a relaxed version of Problem (M). Problem (3) with  $y = y^*$  is a natural way of testing for feasibility in the master problem. From Theorem 2 and the definition of  $V$ ,  $y^*$  satisfies (5) if and only if (3) has a feasible solution for  $y = y^*$ . Also, if (3) turns out to be feasible for  $y = y^*$ , Theorem 3 implies that  $(y^*, r|^*)$  satisfies (10) if and only if  $r \setminus > v(y)$ . With this understanding and by assuming for simplicity that (1) has a finite optimal value, the generalized Benders decomposition **algorithm** is the following:

**Step 1.** Let a point  $y^1$  in  $Y \cap V$  be known. Solve the subproblem (3) for  $y = y^1$  and obtain an optimal (or near-optimal) primal solution  $x^1$  and multiplier vector  $u^1$ . Set  $p = 1, q = 0, UBD = iKy^1$ . Select the convergence tolerance parameter  $\epsilon \succ 0$ .

**Step 2.** Solve the current relaxed master problem:

$$\begin{aligned} & \text{minimize } XJ \\ & \text{Tie \% } y \in Y \end{aligned} \tag{9}$$

s.t.

$$Tj \Rightarrow / (xAy) + (J)^T G(x,y) \quad j = 1, \dots, p \tag{11}$$

$$(A)^T G(xAy) \leq 0 \quad j = 1, \dots, q \tag{12}$$

Let  $(y^*, T|^*)$  be an optimal solution;  $T|^*$  is a lower bound on the optimal value of (1). If  $UBD \leq T|^* + \epsilon$ , terminate.

**Step 3.** Solve the subproblem (3) for  $y = y^*$ . One of the following cases must occur:

**Step 3A.** The quantity  $u(y^*)$  is finite with an optimal primal solution  $x^*$  and an optimal multiplier vector  $u^*$ . If  $i(y^*) \leq r|^* + \epsilon$ , terminate. Otherwise, set  $p = p + 1$ ,

$x^P = x^*$  and  $u^P = u^*$ . If  $v(y^*) < \text{UBD}$ , set  $\text{UBD} = v(y^*)$ . UBD is an upper bound on the optimal value of (1). Return to step 2.

**Step 3B.** Problem  $v(y^*)$  is infeasible for  $y = y^*$ . Determine  $\lambda^*$  in  $\Lambda$  satisfying (13):

$$\left[ \inf_{x \in X} (\lambda^*)^T G(x, y^*) \right] > 0 \quad (13).$$

Let  $x^*$  be the value of  $x$  for which the infimum in (13) is achieved. Set  $q = q + 1$ ,  $x^q = x^*$ ,  $\lambda^q = \lambda^*$ . Return to Step 2.

In the above algorithm we start with a feasible  $y$  point. If such a point is unavailable, Step 1 can be altered to accommodate an infeasible starting point according to Step 3B. The optimization problems in Steps 2 and 3 can be solved by any applicable algorithm. In order to solve (13) in Step 3B, one needs to find a convex combination of the constraints  $G(x, y)$  which are infeasible for subproblem (3). This can be done by any phase-one algorithm for finding an initial feasible solution.

Theoretical convergence of the algorithm in a finite number of steps is guaranteed in the following cases (Geoffrion, 1972):

**Theorem 4 (Finite Convergence):** Assume that  $Y$  is a finite discrete set, that the hypotheses of Theorem 2 hold, and that the hypotheses of Theorem 3 hold with condition (b) omitted. Then, the generalized Benders decomposition procedure terminates in a finite number of steps for any given  $\epsilon > 0$  and even for  $\epsilon = 0$ .

**Theorem 5 (Finite  $\epsilon$ -Convergence):** Assume that  $Y$  is a nonempty compact subset of  $V$ , that  $X$  is a nonempty compact convex set, that  $f$  and  $G$  are convex on  $X$  for each fixed  $y \in Y$  and are continuous on  $X \times Y$ , and that the set  $U(y)$  of optimal multiplier vectors for (3) is nonempty for all  $y$  in  $Y$  and uniformly bounded in some neighborhood of each such point.

Then, for any given  $\epsilon > 0$ , the generalized Benders decomposition procedure terminates in a finite number of steps.

Although the above theorems guarantee finite convergence of the procedure, nothing is said about the number of iterations which are required, or about the point of convergence. The following section discusses the convergence of the algorithm in the special case where the vector of complicating variables belongs to a discrete set. The question of the point of convergence will be addressed in Section 4 where special attention will be given to nonconvex optimization problems.

### 3. Application to Integer Programs

Consider the case of MILP and MINLP models where the vector of complicating variables is binary. The following is then the problem under consideration:

$$\underset{x, y}{\text{minimize}} \quad f(x, y) \quad \text{subject to} \quad G(x, y) \leq 0 \quad y \in \{0, 1\} \quad (14)$$

where for notational simplicity, we have incorporated the constraints  $x \in X$  into  $G(x, y) \leq 0$ . The nonlinear programming relaxation of this MINLP problem is defined as the following NLP:

$$\underset{x, y}{\text{minimize}} \quad f(x, y) \quad \text{subject to} \quad G(x, y) \leq 0 \quad 0 \leq y \leq 1 \quad (15).$$

Geoffrion and Graves (1974), have noticed that reformulating an MILP can have a profound effect upon the efficiency of the GBD algorithm. Subsequently, Magnanti and Wong (1981) addressed the case where the subproblems exhibit several alternative solutions and proposed a methodology for improving the performance of Benders decomposition by finding strong Benders cuts. Among other results, these authors also proved that the convex hull<sup>1</sup>

---

<sup>1</sup> The convex hull formulation of an MILP is defined as the description of the smallest polyhedron (convex space) which includes all the feasible solutions of the MILP. The extreme points of the linear programming relaxation of the convex hull formulation correspond to integer values for the y-variables.

formulation of an MILP requires only one Benders cut in order to converge, namely the cut corresponding to the optimal solution. Here we generalize their result to the larger class of problems where the NLP relaxation in (15) has zero gap<sup>1</sup> with respect to the optimal mixed integer solution of the MINLP in (14):

**Theorem 6:** Assume that the NLP in (15) is a convex problem, *i.e.* that the function  $f(x, y)$  is convex and the constraint set  $G(x, y)$  is also convex in the  $xy$ -space. Further assume that the relaxed problem (15) has an integral (alternative) optimal solution  $(x^*, y^*)$ . Then, when generalized Benders decomposition is applied to the MINLP in (14), termination is obtained in Step 2 after the subproblem in Step 3 is solved at the point  $y = y^*$ .

The proof of this theorem is presented in Appendix A and it makes use of the strong duality and the saddle point theorems for convex NLP problems. The following are some obvious corollaries of Theorem 6:

**Corollary 6.1:** For any LP or convex NLP, Benders decomposition will converge as soon as it determines the optimal solution.

**Corollary 6.2:** For the convex hull formulation of an MILP, Benders decomposition will converge as soon as it determines the optimal solution.

**Corollary 6.3:** For any of the above cases (LP, convex NLP, the convex hull formulation of an MILP, the zero gap formulation of an MINLP), starting with the optimal solution requires only one Benders cut for GBD to converge.

Corollary 6.2 is actually Theorem 4 of Magnanti and Wong (1981). We can see that Theorem 6 above generalizes this result towards two directions. First, it extends to the

---

\* The convex hull formulation is a zero gap formulation but the inverse is not necessarily true. A zero gap formulation has the property that the relaxed problem has the same optimal objective function value as the integer program but the optimal solution of this relaxed problem may exhibit non-integral values for the  $y$ -variables. Certainly, one of the alternative optimal solutions of the relaxed problem is integral.

nonlinear case (MINLP) and, second, it applies to a formulation with zero gap (the convex hull formulation is a special case of the zero gap formulation).

The significance of Corollary 6.3, is that it establishes the fact that starting with the optimal solution of certain formulations, requires only a single iteration for GBD to converge.

To illustrate this consider the following MINLP example:

$$\min \quad f = y_1 + Y_2 + Y_3 + 5x^2 \quad (16)$$

s.t.

$$y_1 + Y_2 \geq 2(1 - Y_3) \quad (17)$$

$$y_1 + Y_2 + Y_3 \geq 2 \quad (18)$$

$$3x \leq y_1 + Y_2 \quad (19)$$

$$0.1Y_2 + 0.25Y_3 \leq x \quad (20)$$

$$0.2 \leq x \leq 1 \quad (21)$$

$$y_t = \{0, 1\} \quad (t=1,2,3) \quad (22)$$

Solving the NLP relaxation of this problem yields an optimal objective function value of  $f = 2.2$  with  $y_1 = 0.6$ ,  $y_2 = 1$ ,  $y_3 = 0.4$ ,  $x = 0.2$ . The problem exhibits a zero gap since the integer programming optimum with  $y_1 = 1$ ,  $y_2 = 1$ ,  $y_3 = 0$ ,  $x = 0.2$  has an optimal objective function value of  $f = 2.2$  as well. In addition, the NLP relaxation is a convex problem and therefore the assumptions of Theorem 6 are satisfied. The results of applying GBD to the MINLP (16)-(22) are shown in Table 1 for two different starting points ( $u_1$  and  $u_2$  are the dual variables corresponding to the constraints (19) and (20) of the LP subproblem). The first starting point is the nonoptimal point  $(y_1, y_2, y_3) = (1, 1, 1)$ . We can see that in this case the procedure converges in 2 iterations right after the optimal integer point  $(y_1, y_2, y_3) = (1, 1, 0)$  is detected. This is in agreement with Theorem 6. In the second case, starting with the

optimum  $(y_1, y_2, y_3) = (1, 1, 0)$  leads to convergence in just one iteration as expected according to Corollary 6.3.

Although as seen in the above example problem and as established by Corollary 6.3, starting with the optimal solution of zero gap formulations requires only a single iteration for GBD to converge, it is well known that if the model at hand does not satisfy the assumptions of Theorem 6, then even starting with the optimal solution may require several Benders iterations for the procedure to build an adequate problem representation in the master problem and to verify the optimality. This observation suggests that the tighter an MINLP formulation is, the fewer the Benders cuts which are required for convergence. This is in accordance with the observations of Geoffrion and Graves (1974) and it points out that, for integer programs, it is important that care be taken in order to develop tight MILP formulations. The development of strong formulations may require developing strong variable upper bounds and using more variables and/or constraints than what standard formulations require. The theoretical issues are discussed elsewhere (eg. Rardin and Choe, 1989; Balas 1985; Jeroslow and Lowe, 1985; Sahinidis and Grossmann 1989 and 1990a). Here we will examine the application of GBD to an example of alternative MILP formulations:

$$\min \quad f = 4 y_1 + y_2 + x_1 + 3 x_2 \quad (23)$$

s.t.

$$x_1 + 2 x_2 = 4 \quad (24)$$

$$x_1 \leq M_1 y_1 \quad (25)$$

$$x_2 \leq M_2 y_2 \quad (26)$$

$$y_i = \{0, 1\} \quad (i=1, 2) \quad (27)$$

where  $M_1$  and  $M_2$  are valid upper bounds for  $x_1$  and  $x_2$ . Specifying different sets of values for  $M_1$  and  $M_2$  gives rise to alternative formulations. Consider the following two cases:

**Case I:**  $M_1 = M_2 = 10$ - The linear programming relaxation of the above MELP has a value of 5.6.

**Case II:**  $M_1 = 4$  and  $M_2 = 2$ . The linear programming relaxation of the above MILP has a value of 7.

Both of the above formulations have the same set of feasible integer solutions and the same optimal solution with  $(y_1, y_2) = (0, 1)$  and an optimal objective function value of 7. Obviously, the second formulation is a zero gap formulation. Therefore, applying GBD to the second formulation and starting with the optimal integer combination  $(y_1 = 0, y_2 = 1)$  leads to convergence in one iteration. On the other hand, as shown in Table 2, from the same starting point the first formulation requires 4 iterations to converge, *i.e.* all integer combinations have to be examined.

#### **4. Application to Nonconvex Optimization**

By making a judicious selection of the complicating variables in the application of generalized Benders decomposition to certain nonconvex nonlinear programming problems, the nonconvexities can be treated separately from the convex portion of the problem by isolating them in the master problem, or they can even be completely eliminated. Geoffrion (1972) was the first one to utilize this principle by applying GBD to the variable factor programming problem. Floudas *et al.* (1989) suggested that the complicating variables be chosen in such a way so as to give rise to convex nonlinear programming subproblems in Step 3 of the algorithm, and to master problems which can be solved to global optimality (eg. linear, convex nonlinear, linear integer programming problems). Even though no mathematical guarantee was given that this approach can identify the global optimum, the procedure was applied to a number of nonconvex problems for which the globally optimal solutions were found from various starting points. On the other hand, however, Sahinidis and Grossmann (1989) have

encountered cases where the application of this technique to their MINLP model for the planning of chemical processes failed to produce the global optima of these problems.

Before we establish the conditions under which the application of GBD provides globally optimal solutions to nonconvex problems, we consider the following example taken from Sahinidis (1990):

$$\min \quad f = -x - y \quad (28)$$

s.t.

$$x \cdot y \leq 4 \quad (29)$$

$$0 \leq x \leq 6 \quad (30)$$

$$0 \leq y \leq 4 \quad (31)$$

The feasible region of this NLP in the  $xy$ -space is shown in Fig. 1. Due to the presence of the nonconvex constraint (29), the problem possesses two strong local minima at points  $(x, y) = (1, 4)$  and  $(x, y) = (6, 2/3)$  with objective function values of  $f = -5$  and  $f = -6.67$ , respectively. Although the second point is the global optimum, the application of any standard NLP optimization code to this problem may lead to the first point which also satisfies the local conditions for optimality.

Let us consider  $y$  as being the complicating variable and apply GBD to this problem. This has the advantage that for a fixed value of  $y$ , the subproblem in Step 3 of the algorithm is an LP which can be solved to global optimality. In addition, the constraints of the master problem have the following form:

$$T_j \rightarrow -x^j - y + u^j (x^j \cdot y - 4)$$

where  $(x^{\wedge}, u^{\wedge})$  is the optimal solution of the LP subproblem at iteration  $i$ . This last constraint is linear and, since  $y$  is a continuous variable, the resulting master problem is again an LP which can be solved to global optimality.

The question whether GBD can locate the global optimum of this problem from any starting point is answered in Table 3 where the results of the application of the algorithm are shown for three different starting points. Starting with  $y^1 = 0$  leads to the local optimum  $y = 4$ ; starting with  $y^1 = 4$  the same local optimum ( $y = 4$ ) is obtained, while starting with  $y^1 = 1.25$  leads to the point  $y = 0.1563$  which is not even a local minimum of the original problem.

In order to understand why the global optimum is not always found, we suggest the following interpretation of the GBD algorithm. The original problem of the form

$$\underset{x, y}{\text{minimize}} \quad f(x, y) \quad \text{subject to} \quad G(x, y) \leq 0 \quad x \in X \quad y \in Y \quad (1)$$

was projected onto the  $y$ -space:

$$\underset{y}{\text{minimize}} \quad \psi(y) \quad \text{subject to} \quad y \in Y \cap V \quad (2)$$

with

$$D(y) = \inf_x [ f(x, y) \quad \text{subject to} \quad G(x, y) \leq 0, \quad x \in X ] \quad (3)$$

$$\text{and} \quad V = \{y: G(x, y) \leq 0 \quad \text{for some } x \in X\} \quad (4)$$

The function  $\psi(y)$  defined in (3) is sometimes called the extremal-value function (Gauvin and Tolle, 1977), the marginal function (Gauvin and Dubeau, 1982), the optimal value function (Fiacco, 1983), or the perturbation function (Rockafellar, 1970; Minoux, 1986). It is a function which arises frequently in parametric nonlinear programming and in the study of optimality conditions of nonlinear optimization problems. Among other properties of

perturbation functions (Chapter 9 of Fiacco, 1983; Chapter 5 of Minoux, 1986), one finds that the lagrangian

$$M \dot{y}) + (\wedge)^T G(\mathbf{A} y)$$

in the right hand side of constraint (11) defines a local support of  $v(y)$  around the point  $y^j$ . Therefore, when the projected problem is convex, this lagrangian produces a valid underestimation of  $i(y)$  for any  $y$  in  $V$  (see Fig. 2 for the case of a linear lagrangian). This fact is exploited in constraint (11) by requiring that the lower bound is greater than this lagrangian function:

$$T_i > l(x^j, y) + (J)^T G(x^j, y) \quad j = 1, \dots, p \quad (11).$$

Generalized Benders decomposition solves the projected problem (2) by building an approximation of  $i(y)$  based on its underestimation (11). For this reason, GBD can be regarded as an outer-approximation algorithm for the solution of the projected problem. Note, however, that unless the projected problem is convex, the local support of the function  $i(y)$  at a point may not provide a valid underestimating function for the entire  $y$ -domain, as shown in Fig. 3 for the case of a linear lagrangian. In this case the GBD procedure will converge in a finite number of steps (according to Theorems 4 and 5), but it may converge to a suboptimal solution since the constraints in (11) may cut off part of the feasible region of the projected problem (2) and the procedure may not provide valid lower bounds.

After this discussion, we can now turn back to our small example problem and investigate the form of the function  $v(y)$ . Since the UP subproblem in Step 3 of the algorithm is a simple problem, we can obtain a closed form solution by analytically solving its optimality conditions. By doing so, it is easy to see that the optimal solution to the LP subproblem is:

$$x \bullet \begin{cases} 6 & \text{if } 0 \leq y \leq 2/3 \\ 4/y & \text{if } 2/3 < y \leq 4 \end{cases}$$

Then the objective function of the projected problem is:

$$v(y) = \begin{cases} -6 - y & \text{if } 0 \leq y \leq 2/3 \\ -4/y - y & \text{if } 2/3 < y \leq 4 \end{cases}$$

This function is not convex as seen in Fig. 4. For this reason, the Benders cuts do not provide valid lower bounds to the objective function and GBD may not identify the global optimum. It can be easily verified (numerically or geometrically by considering the linearizations of  $v(y)$  in Fig. 4) that the point of convergence for this example depends on the starting point used. In particular starting with  $y^1 > 2$  leads to the local minimum  $y = 4$ . Starting with  $y^1 = 2$  leads to the maximum  $y = 2$ . From a starting point in the interval  $(2/3, 2)$  the method will converge to a point in  $[0, 2/3)$ , while starting with  $0 \leq y^1 < 2/3$  leads to the starting point  $y^1$ . It is only with  $y^1 = 2/3$  that the algorithm will converge to the global optimum  $y = 2/3$  !

Having explained why generalized Benders decomposition may fail to identify the global optimum for those cases where the projected problem is not convex, the question which arises is when, if at all, this projected problem is convex. The following result holds when the problem in the  $xy$ -space is convex (Danskin, 1967; Geoffrion, 1971):

**Theorem 7:** Assume that  $X$  and  $Y$  are convex sets, and that  $f$  and each component of  $G$  are convex on  $X \times Y$ . Then the objective  $v(y)$  in (2) is convex on  $Y$ , and the set  $V$  in (4) is convex.

For the case where the original problem is convex or concave, a comprehensive compendium of convexity and concavity properties of the projected problem can be found in Fiacco and Kyparisis (1986). These properties can also be extended to a somewhat wider class of problems which do not satisfy strong convexity or concavity assumptions. This was done by Kyparisis and Fiacco (1987) for several generalized convex (or concave) functions. Nevertheless, since usually a nonconvex NLP does not satisfy any such generalized convexity assumptions, these results cannot be applied in general. Consider as an example the bilinear

program. This is one of the most persistently difficult and recurring nonconvex problems in mathematical programming and it has the following general form:

**Problem BP:**

$$\min \quad J x + y^d A x + d^y$$

s.t.

$$B_x x \leq b_x$$

$$B_y y \leq b_y$$

It would seem that the approach suggested by Floudas *et al.* (1989) is ideal for this type of problem since by any choice of the complicating variables (either  $x$  or  $y$ ) the remaining problem is a simple LP. Furthermore, the master problem for this case is another LP. However, as shown in Appendix B:

**Theorem 8:** The projected problem for the bilinear problem (BP) has a piecewise linear concave objective function.

This theorem implies that the function  $\lambda(y)$  will in general possess several local minima. However, the above result is not to be interpreted that there are no cases for which the application of GBD to the bilinear problem (BP) will lead to the global optimum for any starting point  $y^l$  in  $Y$ . In fact,  $Y$  may be defined in such a way that the restriction of  $\lambda(y)$  over  $Y$  assumes values from only one of the linear segments of  $\lambda(y)$ . In such a case the decomposition approach will locate the global optimum from any starting point as reported by Floudas *et al.* (1989) but note that the original problem in the  $xy$ -space possesses only one local minimizer and therefore any other optimization technique could identify the global solution from any starting point. With this, we come to the main result of this section that if the problem in the  $xy$ -space possesses several local minima, then the projected problem in the  $y$ -space also possess the same number of local minima:

**Theorem 9:** Assume that  $(x_0, y_0)$  is a local minimizer of the original problem in (1) and suppose that:

- (i)  $X$  is a nonempty compact convex set;
- (ii) the functions  $f(x, y)$  and  $G(x, y)$  are continuous on  $X \times V$  and convex in  $X$  for each fixed  $y \in V$ ;
- (iii) there exists a point  $\bar{x} \in X$  such that  $G(\bar{x}, y^*) < 0$  (constraint qualification);
- (iv)  $x_Q$  is the unique optimal solution of the inner optimization problem in (3) for  $y = y_Q$ .

Then,  $(y^*)$  is a local minimizer of the projected problem in (2).

The proof of this theorem is presented in Appendix C and it makes use of properties related to point-to-set mappings and to perturbation functions. Here we would only like to bring to the reader's attention the case of our example problem above where the two local minima of the original problem gave rise to two local minima in the projected problem (at  $y = 2/3$  and  $y = 4$ ).

**Corollary 9.1:** Let the assumptions of Theorem 9 be true for all local minimizers of (1). Then, unless the original problem in (1) possesses a unique local minimum in the  $y$ -space, the projected problem in (2) has multiple local solutions.

For this reason, even though the application of generalized Benders decomposition may result in a very robust and fast procedure which alternates between the solution of two convex problems (subproblem and master problem), there is no guarantee that it will identify the global optima of nonconvex problems. It is then of interest to characterize the points at which the algorithm may converge in this case. It is well known that most NLP algorithms will locate a point that satisfies the Kuhn-Tucker optimality conditions and which is a local optimum for a nonconvex problem. Consider, however, the application of GBD to the nonconvex example of equations (28) to (31). It should be evident from Fig. 4 that any linearizing lagrangian of  $u(y)$

at a point  $2/3 < y < 2$  will cut off both local minima of the problem. Then starting with a point in the interval  $(2/3, 2)$  will lead to convergence to a nonstationary point in the interval  $[0, 2/3)$ . For example, as shown in Table 3, using a starting point of  $y = 1.25$  leads to convergence to the point  $y = 0.1563$ . For this reason, we are led to the following conclusion:

**Remark 9.1:** Although finite convergence of generalized Benders decomposition is guaranteed by Theorem 5, the algorithm in the nonconvex case: (i) may fail to identify the global optimum, (ii) may converge to a point which is not even a local optimum.

**Remark 9.2:** In the convex case the lagrangians provide valid underestimations of the projected function and the algorithm is guaranteed to converge to the global optimum of the problem.

## 6. Conclusions

This paper has discussed issues related to the convergence of generalized Benders decomposition. First, by generalizing a result of Magnanti and Wong, the importance of the tightness of the formulation of an integer program was established. This was achieved by proving that a mixed integer nonlinear programming formulation with zero nonlinear programming relaxation gap requires only the Benders cut of the optimal solution in order to converge. Second, an interpretation of the algorithm as an outer-approximation procedure was given and it was shown that the projected problem on the space of the complicating variables possesses multiple local minima if the original problem also does. For this reason, the application of this technique to nonconvex problems may not always lead to the global optima for these problems; it may not even lead to a local optimum. All the above principles were illustrated through example problems.

## Acknowledgments

The authors would like to acknowledge financial support from the National Science Foundation under Grant CBT-8908735.

## REFERENCES

- Balas, E., "Disjunctive Programming and a Hierarchy of Relaxations for Discrete Optimization Problems", *SIAM J. Alg. Disc. Math.*, **6**(3), 466-486 (1985).
- Bazaraa, M.S., and H.D. Sherali, "Benders' Partitioning Scheme Applied to a New Formulation of the Quadratic Assignment Problem", *Naval Research Logistics Quarterly*, **27**, 29-41 (1980).
- Bazaraa, M.S., and C. M. Shetty, *Nonlinear Programming. Theory and Applications*, John Wiley, New York (1979).
- Benders, J.F., "Partitioning Procedures for Solving Mixed-Variables Programming Problems", *Numerische Mathematik*, **4**, 238-252 (1962).
- Berge, C., *Topological Spaces*, Macmillan, New York (1963).
- Danskin, J.M., *The Theory of Max-Min and its Application to Weapons Allocation Problems*, Springer-Verlag, New York (1967).
- El-Halwagi, M., and V. Manouthisakis, "Design and Analysis of Mass Exchange Networks with Multicomponent Targets", Annual AIChE Meeting, paper 137f, San Francisco (1989).
- Fiacco, A.V., *Introduction to Sensitivity and Stability Analysis of Nonlinear Programming*, Academic Press, New York (1983).
- Fiacco, A.V., and J. Kyparisis, "Convexity and Concavity Properties of the Optimal Value Function in Nonlinear Programming", *JOTA*, **48**(1), 95-126 (1986).
- Florian, M.G., G. Guerin, and G. Bushel, "The Engine Scheduling Problem in a Railway Network", *INFOR J.*, **14**, 121-138 (1976).
- Floudas, C.A., A. Aggarwal, and A.R. Ciric, "Global Optimum Search for Nonconvex NLP and MINLP problems", *Computers and Chemical Engineering*, **13**(10), 1117-1132 (1989).

- Floudas, C.A., and A.R. Ciric, "Strategies for overcoming uncertainties in heat exchanger network synthesis<sup>11</sup>", *Computers and Chemical Engineering*, 13(10), 1133-1152 (1989).
- Gauvin, J., and F. Dubeau, "Differential Properties of the Marginal Function in Mathematical Programming", *Mathematical Programming*, 19, 101-119 (1982).
- Gauvin, J., and F. Tolle, "Differential Stability in Nonlinear Programming", *SIAM J. on Control and Optimization*, 15, 294-311 (1977).
- Geoffrion, A.M., "Elements of Large-Scale Mathematical Programming. Parts I and II.", *Management Science*, 16(11), 652-691 (1970).
- Geoffrion, A.M., "Duality in Nonlinear Programming: A Simplified Applications-Oriented Development", *SIAM Review*, 13(1), 1-37 (1971).
- Geoffrion, A.M., "Generalized Benders Decomposition", *JOTA*, 10(4), 237-260 (1972).
- Geoffrion, A.M., and G.W. Graves, "Multicommodity Distribution System Design by Benders Decomposition", *Management Science*, 20(9), 822-844 (1974).
- Hogan, W., "Point-to-Set Maps in Mathematical Programming", *SIAM Review*, 15(3), 591-603 (1973a).
- Hogan, W., "Directional Derivatives for Extremal-Value Functions with Applications to the Completely Convex Case", *Operations Research*, 21(1), 188-209 (1973b).
- Jeroslow, R.G., and J.K. Lowe, "Experimental Results on the New Techniques for Integer Programming Formulations", *J. Opl. Res. Soc*, 36(5), 393-403 (1985).
- Kyparisis, J., and A.V. Fiacco, "Generalized Convexity and Concavity of the Optimal Value Function in Nonlinear Programming", *Mathematical Programming*, 39, 285-304 (1987).
- Magnanti, T.L., and R.T. Wong, "Accelerating Benders Decomposition: Algorithmic Enhancement and Model Selection Criteria", *Operations Research*, 29(3), 464-484 (1981).

- Minoux, M., *Mathematical Programming: Theory and Algorithms*, Wiley, New York (1986).
- Murty, K.G., *Linear and Combinatorial Programming*, Wiley, New York (1976).
- Noonan, F., and R.J. Giglio, "Planning Electric Power Generation: A Nonlinear Mixed Integer Model Employing Benders Decomposition", *Management Science*, 23(9), 946-956 (1977).
- Rardin, R.L., and U. Choe, *Tighter Relaxations of Fixed Charge Network Flow Problems*, Georgia Institute of Technology, Industrial and Systems Engineering Report Series #J-79-18, Atlanta (1979).
- Rockafellar, R.T., *Convex Analysis*, Princeton University Press, Princeton, NJ (1970).
- Rouhani, R., L. Lasdon, W. Lebow, and A.D. Waren, "A Generalized Benders Decomposition Approach to Reactive Source Planning in Power Systems", *Mathematical Programming Study*, 25,62-75 (1985).
- Sahinidis, N.V., *Mixed Integer Optimization Techniques for Planning and Scheduling of Chemical Processes*, Ph.D. Thesis, Carnegie Mellon University, Pittsburgh, April 1990.
- Sahinidis, N.V., and I.E. Grossmann, "Reformulation of the Multiperiod MELP Model for Capacity Expansion of Chemical Processes", *submitted for publication* (July 1989).
- Sahinidis, N.V., and I.E. Grossmann, "Reformulation of Multiperiod MILP Models for Planning and Scheduling of Chemical Processes", *submitted for publication* (1990a).
- Sahinidis, N.V., and I.E. Grossmann, "MINLP Model for Multiproduct Scheduling on Continuous Parallel Lines", *submitted for publication* (1990b).
- Sahinidis, N.V., I.E. Grossmann, R. Fornari and M. Chathrathi, "Long Range Planning Model for the Chemical Process Industries", *Computers and Chemical Engineering*, 13 (9), 1049-1063 (1989).

**Appendix A: Proof of Theorem 6**

**Theorem 6:** Assume that the NLP in (15) is a convex problem, *i.e.* that the function  $f(x, y)$  is convex and the constraint set  $G(x, y)$  is also convex in the  $xy$ -space. Further assume that the relaxed problem (15) has an integral (alternative) optimal solution  $(x^*, y^*)$ . Then, when generalized Benders decomposition is applied to the MINLP in (14), termination is obtained in Step 2 after the subproblem in Step 3 is solved at the point  $y = y^*$ .

**Proof:** We shall make use of the strong duality and the saddle point theorems for convex NLP problems (see Bazaraa and Shetty, 1979) in order to prove that if the subproblem in Step 3 is solved at the point  $y = y^*$  at iteration  $k$ , then the lower and upper bounds of iteration  $k+1$  satisfy the convergence criterion for  $\epsilon = 0$ .

The Lagrangian dual function for problem (15) is defined as follows:

$$f(x, y, p) = f(x, y) + p^T G(x, y) \quad (\text{A-1})$$

where  $p$  is the vector of dual multipliers. Since (15) is a convex problem, the optimal integer point  $(x^*, y^*, p^*)$  satisfies the strong duality theorem:

$$f(x^*, y^*, p^*) = f(x^*, y^*) \quad (\text{A-2})$$

and from the saddle point theorem we have:

$$f(x^*, y^*, p^*) \leq f(x^*, y, p^*) \quad (\text{A-3})$$

for any  $y \in [0, 1]$  and therefore for any  $y \in \{0, 1\}$ .

Assume now that at iteration  $k$  of the GBD algorithm, the NLP subproblem is solved at the point  $y = y^*$ . This results into an NLP optimal solution of  $x = x^*$  (primal solution) and  $u = p^*$  (dual solution) and therefore an upper bound of

$$\text{UBD} = f(x^*, y^*) \quad (\text{A-4}).$$

The master problem at iteration  $k+1$  is then the following:

$$\underset{\eta \in \mathfrak{R}, y \in \{0, 1\}}{\text{minimize}} \quad \eta \quad (\text{A-5})$$

s. t.

$$\eta \geq f(x^j, y) + (u^j)^T G(x^j, y) \quad j = 1, \dots, p \quad (11)$$

$$(\lambda^j)^T G(x^j, y) \leq 0 \quad j = 1, \dots, q \quad (12)$$

More precisely, the last constraint of (11) reads as follows:

$$\eta \geq f(x^*, y) + (\rho^*)^T G(x^*, y) \quad (11^*)$$

Therefore the solution  $(\hat{y}, \hat{\eta})$  of this master problem will yield a lower bound which is at least equal to the right hand side of (11\*):

$$\begin{aligned} \text{LBD} = \hat{\eta} &\geq f(x^*, \hat{y}) + (\rho^*)^T G(x^*, \hat{y}) && \text{from constraint (11}^*) \\ &= \mathcal{L}(x^*, \hat{y}, \rho^*) && \text{from definition (A-1)} \\ &\geq \mathcal{L}(x^*, y^*, \rho^*) && \text{from the saddle point theorem (A-3)} \\ &= f(x^*, y^*) && \text{from strong duality (A-2)} \\ &= \text{UBD} && \text{from (A-4).} \end{aligned}$$

Therefore the convergence criterion is Step 2 of the procedure will be satisfied with  $\varepsilon = 0$  and the GBD procedure will terminate at this point. ■

**Appendix B: Proof of Theorem 8**

**Theorem 8:** The projected problem for the bilinear problem (BP) has a piecewise linear concave objective function.

**Proof:** By choosing  $y$  as the vector of the complicating variables, the inner problem is to minimize the linear function  $(c^T + y^A) x$  over the linear constraint set  $B_2 x \leq b_v$ . Therefore the projected problem is the inner LP problem parameterized in its objective row coefficients. It is a well known result in linear programming theory that the optimal objective function value in this case is a piecewise linear concave function of the parameters (see Chapter Chapter 7 of Murty, 1976). Therefore  $i(y)$  is a piecewise linear concave function. •

### **Appendix C: Proof of Theorem 9**

We need the following definitions from the theory of point-to-set mappings (Berge, 1963):

**Definition C1:** Given two topological spaces  $Y$  and  $X$ , a point-to-set mapping  $F$  from  $Y$  to  $X$  is a function which associates with every point in  $Y$  a subset of  $X$ .

For example, the function  $\psi(y)$  in (3) defines a point-to-set mapping from  $V$  to  $X$ ; it associates each  $y$  in  $V$  with the subset of  $X$  which is optimal in (3).

**Definition C2:** The point-to-set mapping  $F$  is upper semicontinuous at  $y_0 \in Y$  if, for each open set  $S \subset X$  containing  $F(y_0)$  there exists a neighborhood  $N$  of  $y_0$ ,  $N(y_0)$ , such that for each  $y \in N(y_0)$ ,  $F(y) \subset S$ .

**Definition C3:** The point-to-set mapping  $F$  is upper semicontinuous in  $Y$  if it is upper semicontinuous at each point in  $Y$  and  $F(y)$  is compact for each  $y \in Y$ .

**Definition C4:** The point-to-set mapping  $F$  is lower semicontinuous at  $y_0 \in Y$  if, for each open set  $S \subset X$  satisfying  $S \cap F(y_0) \neq \emptyset$  there exists a neighborhood  $N$  of  $y_0$ ,  $N(y_0)$ , such that for each  $y$  in  $N(y_0)$ ,  $F(y) \cap S \neq \emptyset$ .

**Definition C5:** The point-to-set mapping  $F$  is lower semicontinuous in  $Y$  if it is lower semicontinuous at each point in  $Y$ .

**Definition C6:** The point-to-set mapping  $F$  is said to be continuous in  $Y$  if it is both lower and upper semicontinuous in  $Y$ .

The following is the maximum theorem of Berge (p.116 of Berge, 1963) and refers to the continuity of the perturbation function:

**Theorem C1:** If  $f(x, y)$  is a continuous real-valued function defined on the space  $X \times T$  and  $R(y) = \{x \in X: G(x, y) \leq 0\}$  is a continuous mapping of  $T$  into  $X$  such that  $R(y) \neq \emptyset$  for  $y$  in  $T$ , then the real-valued function defined in (3):

$$v(y) = \inf_x [ f(x, y) \text{ subject to } G(x, y) \leq 0, \quad x \in X ] \quad (3)$$

is continuous in  $T$ . Furthermore, the mapping  $F$ , defined by

$$F(y) = \{x \in R(y): f(x, y) = v(y)\} \quad (C-1)$$

is an upper semicontinuous mapping of  $T$  into  $X$ .

An alternative to the Berge definition of continuity of a map is given in the setting of open and closed maps by Hogan (1973a). The notions of lower and upper semicontinuity are equivalent to the notions of open and closed, respectively, provided that the set  $X$  is compact (Hogan, 1973a). Based on these definitions, Hogan (1973b) has derived the following result concerning the continuity of the map  $R(y)$ :

**Theorem C2:** Let the functions  $f(x, y)$  and  $G(x, y)$  be continuous on  $X \times V$  and let the functions  $G(x, y)$  be convex for each fixed  $y \in V$ . Also assume that  $X$  is closed and convex, and that for each  $y \in V$  there exists a point  $\bar{x} \in X$  such that  $G(\bar{x}, y) < 0$ . Then  $R(y) = \{x \in X: G(x, y) \leq 0\}$  is a continuous mapping in  $V$ .

With this background we can proceed to the proof.

**Theorem 9:** Assume that  $(x_0, y_0)$  is a local minimizer of the original problem in (1) and suppose that:

- (i)  $X$  is a nonempty compact convex set;
- (ii) the functions  $f(x, y)$  and  $G(x, y)$  are continuous on  $X \times V$  and convex in  $X$  for each fixed  $y \in V$ ;
- (iii) there exists a point  $\bar{x} \in X$  such that  $G(\bar{x}, y_0) < 0$  (constraint qualification);

(iv)  $x_0$  is the unique optimal solution of the inner optimization problem in (3) for  $y = y_0$ .

Then,  $(y_0)$  is a local minimizer of the projected problem in (2).

**Proof:** Since  $(x_0, y_0)$  is a local minimizer of (1), there exist balls  $B_{x_0} = \{x \in X: \|x - x_0\| \leq \delta_{x_0} \text{ with } \delta_{x_0} > 0\}$  and  $B_{y_0} = \{y \in Y: \|y - y_0\| \leq \delta_{y_0} \text{ with } \delta_{y_0} > 0\}$  around  $x_0$  and  $y_0$  such that

$$f(x_0, y_0) \leq f(x, y) \quad \text{for any } (x, y) \in (B_{x_0} \times B_{y_0}) \cap W \quad (\text{C-2})$$

where  $W = \{(x, y) \in X \times V: G(x, y) \leq 0\}$ .

In order to show that  $y_0$  is a local minimizer of (2), we need to show that  $v(y_0) \leq v(y)$  for  $y$  in a small neighborhood around  $y_0$ , i.e. for  $y \in B_y(y_0) = \{y \in V \cap Y: \|y - y_0\| \leq \varepsilon_{y_0} \text{ with } \varepsilon_{y_0} > 0\}$ .

The assumptions of Theorem C2 are satisfied and therefore the mapping  $R(y)$  is continuous in  $V$ . Since by the definition of  $V$  in (4),  $R(y) \neq \emptyset$  for  $y$  in  $V$ , we can apply Theorem C1 with  $V$  in place of  $T$ . From this we conclude that the mapping  $\Gamma(y)$  defined by (C-1) is upper semicontinuous in  $V$ . Let  $S = \{x \in X: \|x - x_0\| < \delta_{x_0}\}$  be the open subset of  $B_{x_0}$  containing  $\Gamma(y_0) = \{x_0\}$ . Then from the definition of the upper semicontinuity of the point-to-set mapping  $\Gamma(y_0)$ , it follows that there exists a neighborhood  $N(y_0)$  in  $V$  such that for any  $y$  in  $N(y_0)$  we have  $\Gamma(y) \subset S \subset B_{x_0}$ . Let  $B_y(y_0)$  be the intersection of  $N(y_0)$  and  $B_{y_0}$ ; that is  $B_y(y_0) = \{y \in V \cap Y: \|y - y_0\| \leq \delta_{y_0}\}$ . This is a nonempty set since it contains the point  $y_0$ . Then for any point  $y^*$  in  $B_y(y_0)$  we have  $\Gamma(y^*) \subset B_{x_0}$ . Since for any point  $(x^*, y^*) \in [\Gamma(y^*) \times B_y(y_0)] \cap W$  we have  $(x^*, y^*) \in (B_{x_0} \times B_{y_0}) \cap W$ , it follows from (C-2) that  $v(y^*) = f(x^*, y^*) \geq f(x_0, y_0)$ . Therefore  $y_0$  satisfies the local optimality conditions of (2) in the neighborhood defined by  $B_y(y_0)$ . ■

**Table 1.** Application of GBD to zero gap MINLP example problem.

**Starting Point: nonoptimal**

Iteration (/)	$(y_1, y_2, y_3)$	$y^*$	$(u_1, u_2)^j$	UBD	LBD
1	(1, 1, 1)	0.35	<b>(0, 3.5)</b>	3.6125	1.7375
2	(1, 1, 0)	<b>0.2</b>	<b>(0,0)</b>	<b>2.2</b>	<b>2.2</b>

**Starting Point: optimal**

Iteration (/)	$(y_1, y_2, y^*)$	$y^*$	$(u_1, u_2)^j$	UBD	LBD
1	(1, 1, 0)	<b>0.2</b>	<b>(0,0)</b>	<b>2.2</b>	<b>2.2</b>

**Table 2.** Application of Benders decomposition to weak MILP formulation.

Iteration ( $l$ )	$(y_1, y^2)$	$(x_1, x_2)$	$(u_1, u_2)$	UBD	LBD
1	(0,1)	(0,2)	(1,0)	7	0
2	(1,0)	(4,0)	(0,1)	7	1
3	(1,1)	(4,0)	(0,0)	7	6
4	(0,0)	$(\frac{1}{3}, \frac{1}{3})$	$(\frac{1}{3}, \frac{1}{3})$	7	7

**Table 3.** Application of GBD to nonconvex NLP example problem.

**Starting Point:  $y = 0$**

Iteration ( $j$ )	$y^j$	$x^j$	$w^j$	UBD	LBD
1	0	6	0	-6	-10
2	4	1	$\frac{1}{4}$	-6	-5

**Starting Point:  $y = 4$**

Iteration ( $j$ )	$y^j$	$x^j$	$w^j$	UBD	LBD
1	4	1	$\frac{1}{4}$	-5	-5

**Starting Point:  $y = 1.25$**

Iteration ( $j$ )	$y^j$	$x^j$	$w^j$	UBD	LBD
1	1.25	3.2	0.8	-4.45	-6.4
2	0	6	0	-6	-6.1563
3	0.1563	6	0	-6.1563	

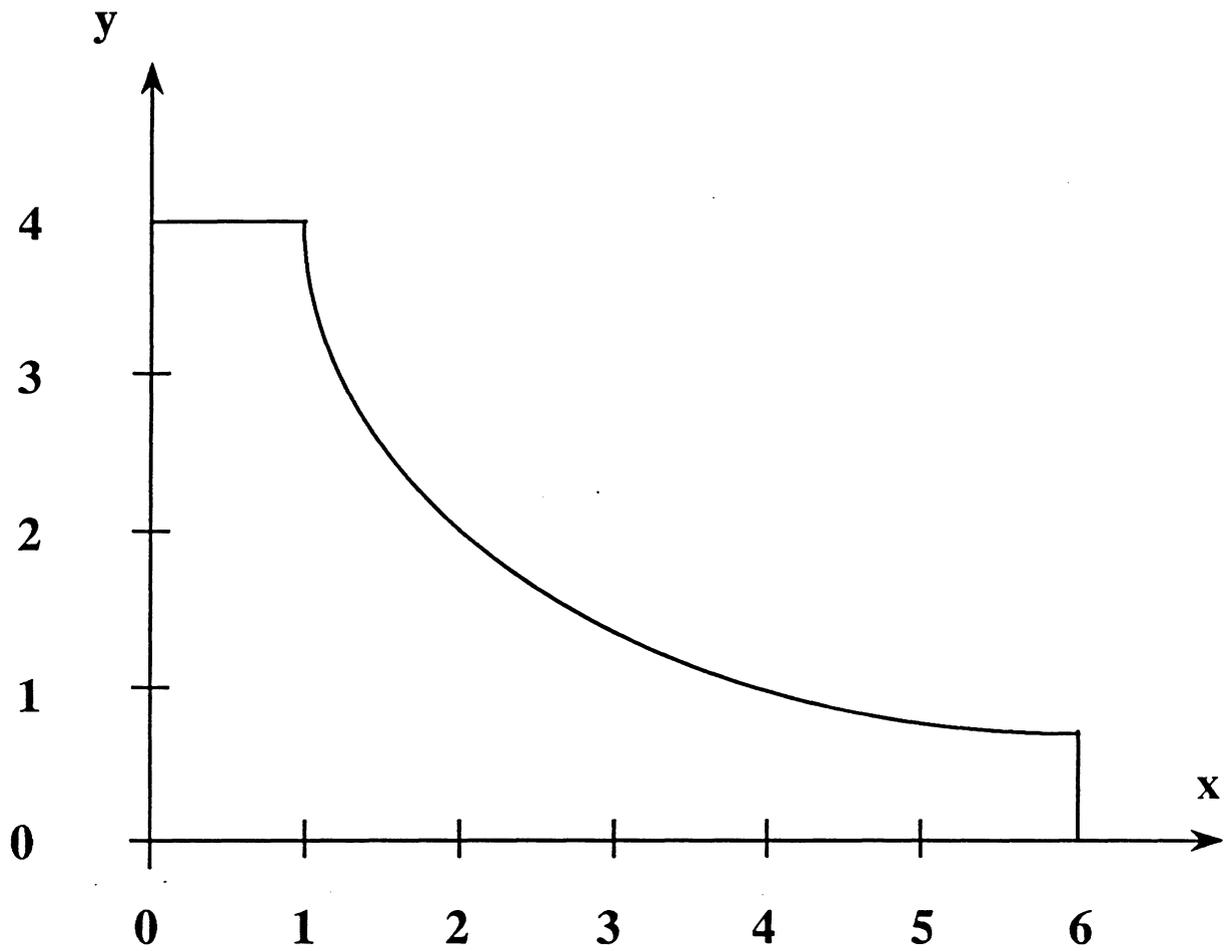
## **LIST OF FIGURES**

**Fig. 1:** Feasible region of nonconvex  
NLP example problem.

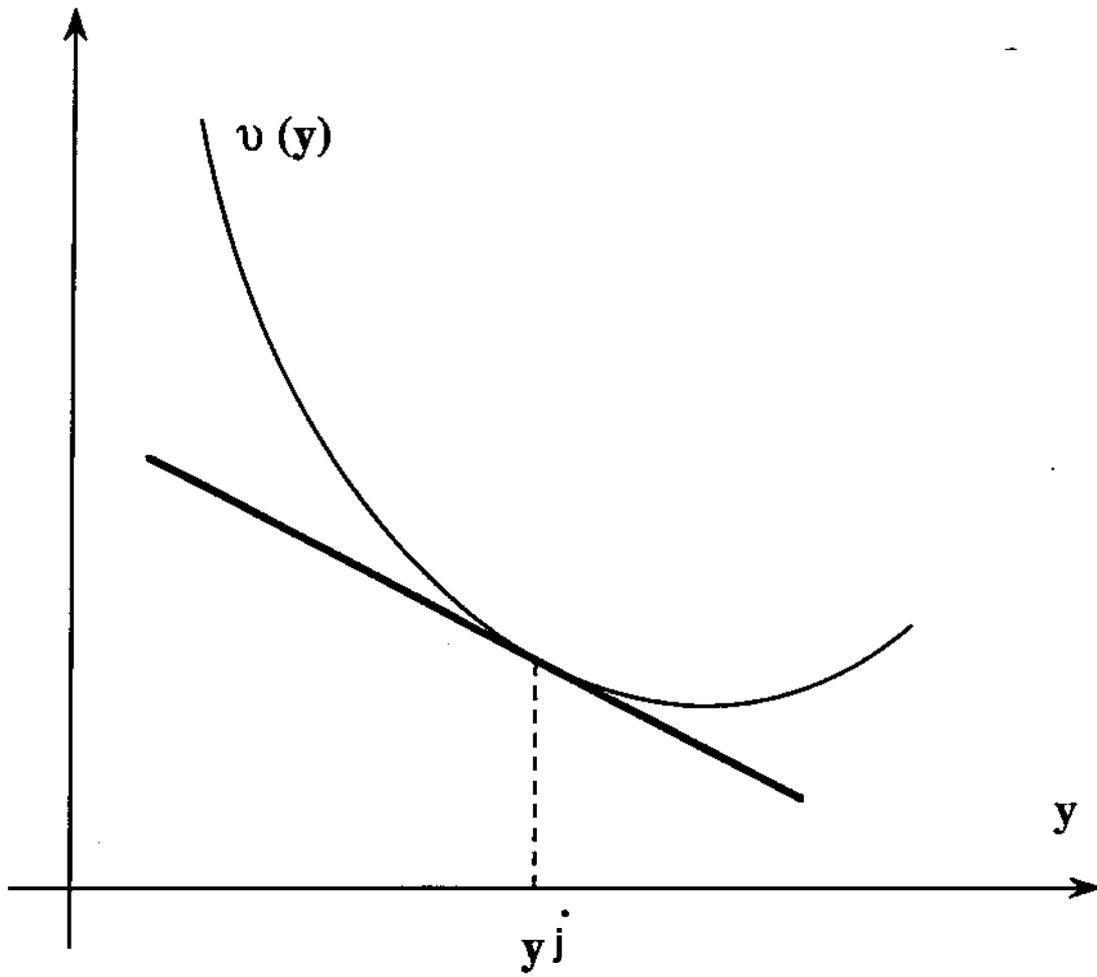
**Fig. 2:** Linear lagrangian support  
for a convex function.

**Fig. 3:** Linear lagrangian support  
for a nonconvex function.

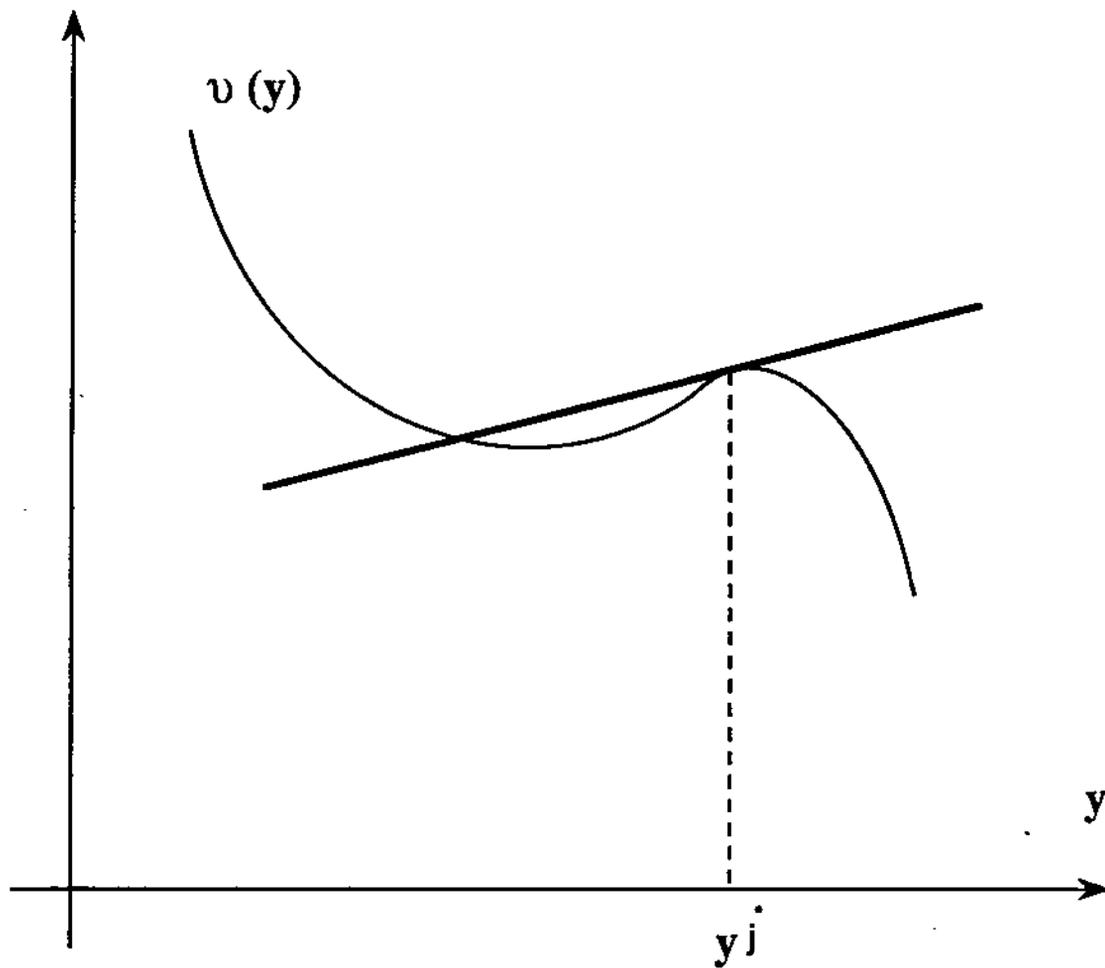
**Fig. 4:** Objective function in projected space.



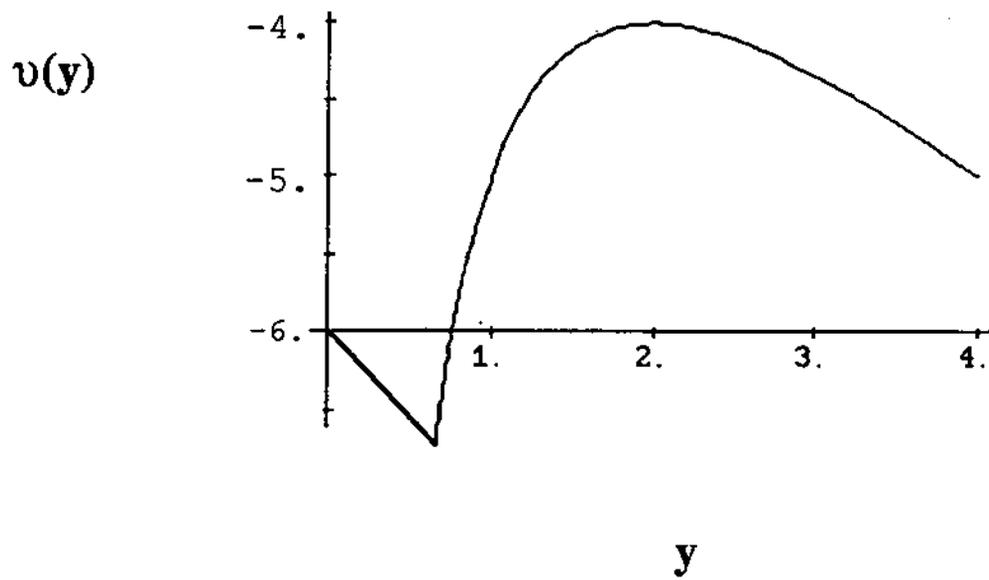
**Fig. 1: Feasible region of nonconvex  
NLP example problem**



**Fig. 2:** Linear Lagrangian support  
for a convex function



**Fig. 3:** Linear lagrangian support  
for a nonconvex function



**Fig. 4:** Objective function in projected space