Computers chem. Engng, Vol. 15, No. 10, pp. 691–700, 1991
Printed in Great Britain.

0098-1354/91 $3.00 + 0.00

ON THE GENERALIZED BENDERS DECOMPOSITION

M. J. Bagajewicz and V. Manousiouthakis†
Department of Chemical Engineering, UCLA, Los Angeles, CA 90024, U.S.A.

(*Received 27 August 1990; final revision received 26 June 1991; received for publication 11 July 1991*)

Abstract—Generalized Benders Decomposition is a procedure to solve certain types of NLP and MINLP problems. The use of this procedure has been recently suggested as a tool for solving process design problems. This paper analyzes the solution of nonconvex problems through different implementations of the Generalized Benders Decomposition. It is demonstrated that in certain cases only local minima may be found, whereas in other cases not even convergence to local optima can be achieved. A criterion to identify whether the converged value is a candidate for being a local minimum is provided. It is also shown that in the presence of a dual gap, a particular implementation of the Generalized Benders Decomposition may provide upper and lower bounds on the global optimum.

1. INTRODUCTION

Generalized Benders Decomposition has been suggested as a solution procedure for certain NLP and MINLP problems (Geoffrion, 1972). Its name is coined after the decomposition procedure first developed by Benders for the solution of mixed-variable programming problems (Benders, 1962). However, some restrictions regarding the convexity and other properties of the functions involved were identified. A recent paper (Floudas *et al.*, 1989) reviews this technique revealing its potential for application in chemical process design. It also proposes a computational implementation that is claimed to potentially identify global optima in nonconvex NLP and MINLP problems. This claim is sustained by solving different examples, though not proven mathematically.

In this paper we point out the importance of two properties defined by Geoffrion (1972), namely *Property* (*P*) and *L-Dual-Adequacy*, for proper application of the technique. In Section 2, we review Generalized Benders Decomposition. In Section 3.5, we show that the implementation of Benders iterations when *Property* (*P*) is not satisfied, can identify as solutions, points that are not even local extrema. In Section 3.6 we show that not satisfying the requirement of *L-Dual-Adequacy* can also lead to the identification of local minima or points that are not even extrema. Moreover, it is shown that when these properties are assumed to be satisfied (although they may not hold) and the starting points are local minima or maxima, then the algorithm termination criterion is reached at the first iteration. In Section 3.7, the local character of the implementations illustrated in Sections 3.2 and 3.6 is discussed.

Although the implementation of Generalized Benders Decomposition does not identify global extrema when these properties are not satisfied, a criterion to determine if the convergence point is a candidate for being a local extremum is provided in Section 3.8. Thus, with the help of this criterion we show that Generalized Benders Decomposition may be used even when *Property* (*P*) and/or *L-Dual-Adequacy* are violated, to identify local extrema.

Finally, in Section 4, an implementation of the Generalized Benders Decomposition that satisfies these two crucial properties, is presented. Through examples and mathematical analysis it is shown that dual gaps may prevent this procedure from converging to the global solution, and therefore only bounds on the global optimum may be obtained.

2. GENERALIZED BENDERS DECOMPOSITION

Consider the optimization problem:

$$\begin{aligned} \min_{x,y}\ & F(x, y), \\ \text{s.t.}\ & G(x, y) \leqslant 0, \\ & x \in X, \\ & y \in Y. \end{aligned} \tag{1}$$

The following problem is equivalent to (1) and is called its projection on Y (Geoffrion, 1972):

$$\begin{aligned} \min_{y}\ & v(y), \\ \text{s.t.}\ & v(y) = \begin{Bmatrix} \min_{x \in X} & F(x, y) \\ \text{s.t.} & G(x, y) \leqslant 0 \end{Bmatrix} \\ & y \in Y \cap V, \end{aligned} \tag{2}$$

where $V = \{y: G(x, y) \leqslant 0 \textit{ for some } x \in X\}$.

For problems where X is a convex set, and the functions $F(x, y)$, $G(x, y)$ are convex with respect to

†To whom all correspondence should be addressed.

the variable x, Geoffrion (1972) proposed a decomposition of (2) based on the following two problems:

Primal problem:

$$\min_{x \in X} F(x, \bar{y}),$$
$$\text{s.t.} \quad G(x, \bar{y}) \leqslant 0,$$
$$x \in X, \qquad (3)$$

where $\bar{y}$ is an arbitrary but fixed point in Y.

Master problem:

$$\min_{y \in Y} [\max_{u \geqslant 0} [\min_{x \in X} \{F(x, y) + u^T G(x, y)\}]],$$
$$\text{s.t.} \min_{x \in X} \{\lambda^T G(x, y)\} \leqslant 0, \quad \forall \lambda \in \Lambda, \qquad (4)$$

where $\Lambda = \{\lambda \geqslant 0; \Sigma_i \lambda_i = 1\}$

Remark 2.1

Before any further analysis, it is appropriate to point out that the master problem is equivalent to the projection (2), only when X is a convex set, and the functions $F(x, y)$, $G(x, y)$ are convex with respect to x. This is so because the dual (Luenberger, 1969; Rockafellar, 1972) of $v(y)$, as defined in (2), was invoked in arriving at (4). Therefore, when X is not a convex set and/or convexity of the functions $F(x, y)$ and $G(x, y)$ in the variable x does not hold, a dual gap may exist between $v(y)$ and its dual. In these cases, since by the weak duality theorem, the solution of $v(y)$ is always greater than or equal to the solution of its dual, (Luenberger, p. 225) the master problem can only provide a lower bound on the optimum, i.e.

$$v(y) = \begin{Bmatrix} \min_{x \in X} & F(x, y) \\ \text{s.t.} & G(x, y) \leqslant 0 \end{Bmatrix}$$
$$\geqslant \max_{u \geqslant 0} [\min_{x \in X} \{F(x, y) + u^T G(x, y)\}]$$
$$\Rightarrow$$
$$\min_{y \in Y \cap V} v(y) \geqslant \min_{y \in Y \cap V} [\max_{u \geqslant 0} [\min_{x \in X} \{F(x, y) + u^T G(x, y)\}]]$$
$$= \begin{Bmatrix} \min_{y \in Y} [\max_{u \geqslant 0} [\min_{x \in X} \{F(x, y) + u^T G(x, y)\}]] \\ \text{s.t.} \min_{x \in X} \{\lambda^T G(x, y)\} \leqslant 0, \quad \forall \lambda \in \Lambda \end{Bmatrix}.$$

Remark 2.2

The constraint of the master problem $\min_{x \in X}\{\lambda^T G(x, y)\} \leqslant 0$ serves the purpose of generating solutions $\hat{y}$ that are in the set V. For the convex case, this has been shown by Geoffrion (1972) in Theorem 2.2 which employs a corollary of Theorem 5 (Geoffrion, 1971). If $G(x, y)$ is not convex in x and/or X is not a convex set, the theorem is still valid, as discussed in the Appendix.

2.1. Computational implementation

The master problem can be rewritten as:

$$\min_{y_0, y} y_0$$
$$\text{s.t.} \quad L^*(y, u) \triangleq \min_{x \in X} \{F(x, y) + u^T G(x, y)\} \leqslant y_0,$$
$$\text{all } u \geqslant 0, \qquad (5)$$
$$L_*(y, \lambda) \triangleq \min_{x \in X} \{\lambda^T G(x, y)\} \leqslant 0,$$
$$\text{all } \lambda \in \Lambda. \qquad (6)$$

Remark 2.3

Geoffrion (1972) suggests to solve a relaxed version of this problem in which all but a few constraints are ignored. Since constraints are continuously added to the master, the optimal values of this problem form a monotone nondecreasing sequence. When $F(x, y)$ and $G(x, y)$ are convex in x and certain conditions hold (Geoffrion, 1972, Theorem 2.5), the lowest solution of the primal and the global solution of the master [as defined by (5) and (6)] will approach one another and will thus provide the global optimum of the overall problem within a prespecified tolerance. As analyzed earlier (Remark 2.1), when convexity in x does not hold, dual gaps may prevent these two values from approaching each other. Nevertheless the global solution of the relaxed master problem will still provide a valid lower bound on the global optimum of the overall problem. The resulting computational procedure is the following:

Step 1. Let a point $\bar{y}$ in $Y \cap V$ be known. Solve the primal problem and obtain the optimal solution x^*, and the optimal multiplier vector u^*. Put the counters $K^f = 1$, $K^i = 0$. Set $U = F(x^*, \bar{y})$, select a tolerance $\epsilon > 0$ and put $u^{(1)} = u^*$. Finally, determine the function $L^*(y, u^{(1)})$.

Step 2. Solve globally the current relaxed master problem:

$$\min_{y_0, y} y_0$$
$$\text{s.t.} \quad L^*(y, u^{(k_1)}) \leqslant y_0 \quad k_1 = 1, \ldots K^f,$$
$$L_*(y, \lambda^{(k_2)}) \leqslant 0 \quad k_2 = 1, \ldots K^i.$$

Let $(\hat{y}, \hat{y}_0)$ be the *globally* optimal solution. The optimal value $\hat{y}_0$ is a lower bound on the optimal value of the global problem. If $U \leqslant \hat{y}_0 + \epsilon$, terminate.

Step 3. Solve globally the primal problem using $\bar{y} = \hat{y}$.

Step 3a. *Primal is feasible*. If $v(\bar{y}) \leqslant \hat{y}_0 + \epsilon$ terminate. Otherwise, determine the optimal multiplier vector u^*, increase K^f by one, set $u^{(K^f)} = u^*$. Additionally, if $v(\bar{y}) < U$, put $U = v(\bar{y})$. Finally, determine the function $L^*(y, u^{(K^f)})$ and return to Step 2.

Step 3b. *Primal is infeasible.* Determine a set of values of $\lambda^* \in \Lambda$ which satisfy

$$\min_{x \in X} \{\lambda^{*T} G(x, y)\} > 0.$$

Increase K^i by one, put $\lambda^{(K^i)} = \lambda^*$ and determine the function $L_*(y, \lambda^{(K^i)})$. Return to Step 2.

Remark 2.4

When referring to "solutions" of nonconvex optimization problems it is sometimes understood that these may be locally, rather than globally, optimal points. In the above computational procedure it is necessary that the master be solved globally. When problems are jointly convex in x and y, the master is convex (Geoffrion, 1972, Section 4.2) and globality is then achieved. Global solutions of the primal are also needed if convexity of X and/or convexity of $F(x, y)$ and/or $G(x, y)$ in x does not hold.

2.2. Determination of λ^*

The determination of λ^* in Step 3b can be done by any Phase I algorithm. In particular Floudas *et al.* (1989) proposed to solve the following problem:

$$\begin{aligned} &\min_{\alpha, x} \alpha \\ &\text{s.t.} \quad G(x, \bar{y}) - \alpha \underline{1} \leq 0, \\ &\qquad x \in X, \\ &\qquad \alpha \in R, \end{aligned} \tag{7}$$

where $\underline{1} = (1, 1, \ldots, 1)^T$.

Since the primal problem is infeasible, we already know that the solution of this problem is positive. Once this problem is solved and a stationary point x^* is obtained, the following necessary Kuhn–Tucker conditions are satisfied:

$$\begin{aligned} 1 - \sum_i \lambda_i^* &= 0, \\ \lambda^{*T} \nabla_x G(x^*, \bar{y}) &= 0, \\ \lambda_i^* [G_i(x^*, \bar{y}) - \alpha] &= 0, \quad \forall i, \\ \lambda_i &\geq 0, \quad \forall i. \end{aligned} \tag{8}$$

From them we conclude that by solving this problem we satisfy the condition $\lambda^* \in \Lambda$. Note that in this step it is not imperative to achieve globality.

2.3. Explicit determination of $L^*(y, u^*)$ and $L_*(y, \lambda^*)$

In most cases, the functions L^* and L_* are implicitly defined. *One case* in which these functions can be obtained in explicit form is when the global minimum over x can be obtained independently of y, i.e. when *Property* (P) is satisfied. Two examples in which this property is satisfied arise when the functions $F(x, y)$ and $G(x, y)$ are separable in x and y, and in the variable factor programming case.

Once *Property* (P) holds, evaluation of $L^*(y, u^*)$, $L_*(y, u^*)$ simply requires that the minima in (5) and (6) be global. This is the so called *L-Dual-Adequacy* property, which upon satisfaction of *Property* (P) can be achieved only if a global search of the solutions of both

$$\min_{x \in X} \{F(x, y) + u^{*T} G(x, y)\} \quad \text{and} \quad \min_{x \in X} \{\lambda^{*T} G(x, y)\}$$

is conducted.

Another case in which these functions can be obtained in explicit form is when *Property* (P') (Geoffrion, 1972) is satisfied, i.e. when the globally optimal solution of the primal, x^*, is also the solution of the minimization problems defined in (5–6) (for all y). This is guaranteed when $F(x, y)$ and $G(x, y)$ are convex and separable in x.

If *Property* (P') is simply assumed (Floudas *et al.*, 1989), then this implementation of Generalized Benders Decomposition may only be used, without guarantees, to identify candidates for local, but not global minima.

The above discussion does not mean that *Properties* (P) and/or (P') must always hold. Other procedures may exist in which the Master problem [as defined by (5) and (6)] can be solved without using these requirements.

3. NP-GENERALIZED BENDERS ALGORITHM

Let us call the version of the Generalized Benders Algorithm in which *Property* (P') is simply assumed, *NP-Generalized Benders Algorithm* (NP-GBA).

Specifically, in the NP-GBA, x^* are used as proposed "solutions" of the minimization problem involving the determination of $L^*(y, u^*)$ or $L_*(y, \lambda^*)$. With this done, the master problem can be rewritten as follows:

$$\begin{aligned} &\min_{y_0, y} y_0 \\ &\text{s.t.} \quad F(x^{(k_1)}, y) + u^{(k_1)T} G(x^{(k_1)}, y) \leq y_0, \quad k_1 = 1, \ldots K^f, \\ &\qquad \lambda^{(k_2)T} G(x^{(k_2)}, y) \leq 0, \quad k_2 = 1, \ldots K^i, \end{aligned}$$

where $x^{(k_1)}$ and $x^{(k_2)}$ are the respective solutions of the primal and $u^{(k_1)}$, $\lambda^{(k_2)}$ are the corresponding multipliers.

3.1. The termination and convergence criteria

As proposed by Geoffrion (1972), we say that the Generalized Benders Algorithm's convergence criterion is:

> If the optimal value of the master problem approaches from below the upper bound provided by the primal, within a convergence error ϵ, then the GBD algorithm stops with the upper bound representing the optimal solution.

Under this implementation of the Generalized Benders Decomposition, the solutions of the master problem may exceed the values given by the primal problem. This happens because *Property* (P) and/or

L-Dual-Adequacy do not hold. Therefore a *termination criterion* is added to account for these situations. The new termination criterion is the following:

If the optimal value of the master problem, $\hat{y}_0$, exceeds the current upper bound then the algorithm stops with the upper bound representing the optimal solution (Floudas *et al.*, 1989).

When the termination criterion is not activated and the relaxed master sequence of the NP-GBA converges, then the master's converged value is equal to the upper bound obtained from the primal's solution. This is true even for nonconvex problems which may exhibit a gap between $v(y)$ and its dual. Indeed, upon convergence of the relaxed master sequence, the Kuhn–Tucker conditions of the corresponding primal suggest that $u^{(k_1)T}G(x^{(k_1)}, y) = 0$. In turn, this implies that one of the master's constraints becomes $y_0 \geq F(x^{(k_1)}, y)$. The termination criterion is not activated and thus $y_0 \geq F(x^{(k_1)}, y)$ will render $y_0 = F(x^{(k_1)}, y)$.

3.2. The use of NP-GBA when Property (P) does not hold

We will present now an optimization problem that will serve to demonstrate the anomalies of the NP–GBA.

Example

Consider the following optimization problem:

$$\begin{aligned} \min_{z_1, z_2} \quad & F(z_1, z_2) = f(z_1) - z_2, \\ \text{s.t.} \quad & z_2 - g(z_1) \leq 0, \\ & z_1 + z_2 \leq a, \\ & z_1 \geq 0.9, \end{aligned}$$

where $a = 6403/150$ and:

$$f(z_1) = 4z_1^5 - \tfrac{45}{2}z_1^4 + \tfrac{130}{3}z_1^3 - 18z_1^2 - \tfrac{4}{3}z_1,$$

$$g(z_1) = 12z_1^2 - \tfrac{4}{3}z_1.$$

The problem is nonconvex and has two local minima, one at $(z_1, z_2) = (1, 10.666)$ and the other at $(z_1, z_2) = (1.9, 3059/75)$, with corresponding values of the objective function (-5.16666) and (-5.25495), respectively. It also has two local maxima, located at $(z_1, z_2) = (1.5, 25)$ with an objective function value of -4.78125 and at $(z_1, z_2) = (0.9, 8.52)$ with an objective function value of -5.11029.

In the spirit of the approach suggested by Floudas *et al.* (1989), we introduce the new variable vector $w = [w_1 \ w_2 \ w_3 \ w_4]^T$, so that the problem is rewritten in the following form:

$$\begin{aligned} \min_{w, z_2, y} \quad & F_w(w, z_2, y) = M(w)y - z_2, \\ \text{s.t.} \quad & q(y, z_2, w) = z_2 - 12w_4 y + \tfrac{4}{3}y \leq 0, \\ & r_1(y, w) = w_1 - w_2 y = 0, \\ & r_2(y, w) = w_2 - w_3 y = 0, \\ & r_3(y, w) = w_3 - w_4 y = 0, \\ & r_4(y, w) = w_4 - y = 0, \\ & s(y, z_2) = z_2 + y - a \leq 0, \\ & y \geq 0.9, \\ & (w, z_2) \in X = R^5, \\ & y \in Y = R, \end{aligned}$$

where:

$$M(w) = 4w_1 - \tfrac{45}{2}w_2 + \tfrac{130}{3}w_3 - 18w_4 - \tfrac{4}{3}.$$

In this alternative formulation this problem is nonconvex, but the introduction of additional variables has made it bilinear. Therefore the choice of y as a complicating variable will render linear the NP-GBA's primal and master problems.

The primal problem becomes:

$$\begin{aligned} \min_{w, z_2, y} \quad & F_w(w, z_2, \bar{y}) = M(w)\bar{y} - z_2, \\ \text{s.t.} \quad & q(\bar{y}, w, z_2) \leq 0, \\ & r_1(\bar{y}, w) = 0, \\ & r_2(\bar{y}, w) = 0, \\ & r_3(\bar{y}, w) = 0, \\ & r_4(\bar{y}, w) = 0, \\ & s(\bar{y}, z_2) \leq 0, \\ & (w, z_2) \in X = R^5. \end{aligned}$$

Using NP-GBA, the master problem becomes:

$$\begin{aligned} \min_{y_0, y} \quad & y_0 \\ \text{s.t.} \quad & M(w^*)y - z_2^* + u_q^* q(y, z_2^*, w^*) \\ & \quad + \sum_{i=1}^{4} u_{r_i}^* r_i(y, w^*) + u_s^* s(y, z_2^*) \\ & \quad \triangleq A^* + B^* y \leq y_0, \\ & \lambda_q^* q(y, z_2^*, w^*) + \sum_{i=1}^{4} \lambda_{r_i}^* r_i(y, w^*) + \lambda_s^* s(y, z_2^*) \\ & \quad \triangleq A_* + B_* y \leq 0, \\ & y \geq 0.9, \quad (y_0, y) \in R^2. \end{aligned}$$

All of the requirements of the Generalized Benders Decomposition procedure hold, i.e. $F(x, y)$ and $G(x, y)$ are convex in x, except that *Property (P)* is not satisfied. By solving this problem using NP-GBA we will demonstrate the following anomalies:

—Convergence to a local minimum.
—Termination achieved when starting near or at a local maximum.
—Convergence to nonextremum points.

3.3. *Convergence to a local minimum*

First note that if the starting point is at the local minimum $\bar{y} = 1$, then the optimal solution is $(w_1^*, w_2^*, w_3^*, w_4^*, z_2^*) = (1, 1, 1, 1, 10.6666)$ with an optimal value of -5.1666. The optimal multipliers are $(u_q^*, u_{r_1}^*, u_{r_2}^*, u_{r_3}^*, u_{r_4}^*, u_s^*) = (1, -4, 18.5, -24.8333, 5.1666, 0)$ and the NP-GBA's master problem becomes:

$$\min_{y_0, y} y_0,$$

$$\text{s.t.} \quad -5.1666 \leqslant y_0,$$

which has an optimal value of $y_0 = -5.1666$. Therefore, the termination criterion is met.

The procedure has identified an "optimum" at $(y, x) = (1, 10.666)$. Starting at the other local optima with $\bar{y} = 1.9$ the terminating criterion is also immediately met in the same way.

If for example a point near the local minimum, $\bar{y} = 1.1$, is chosen as a starting point, the algorithm converges again to the same local minimum (1, 10.666). The primal problem is always feasible and relevant results for each iteration are shown in Table 1.

Note that at the second iteration, the master provides a value which is already greater than the global optimum.

3.4. *Termination achieved when starting near or at a local maximum*

At this point one may suggest that although the procedure will not identify global optima, it can be used as a good local optima finder. However, this statement should also be used with caution because the termination criterion can be met if the starting point is a local maximum.

Indeed, starting at $\bar{y} = 1.5$, which is a local maximum of the problem, one obtains the following solution of the primal problem $(w_1^*, w_2^*, w_3^*, w_4^*, z_2) = (5.0625, 3.375, 2.25, 1.5, 25.)$ with an optimal value of -4.78125. The optimal multipliers are: $(u_q^*, u_{r_1}^*, u_{r_2}^*, u_{r_3}^*, u_{r_4}^*, u_s^*) = (1, -6, 24.75, -27.875, 3.1875, 0)$.

The NP-GBA's master problem becomes:

$$\min_{y_0, y} y_0,$$

$$\text{s.t.} \quad -4.78125 \leqslant y_0.$$

which has an optimal value of $y_0 = -4.78125$. Thus, according to the termination criterion, the "optimum" has been identified.

Additionally, if one analyzes the behavior of the NP-GBD at other starting points, the termination criterion may also be met. Take for example $\bar{y} = 1.49$ as a starting point. The first two iterations will provide a value for the master problem which is higher than the current updated upper bound, so that the termination criterion is met. This is illustrated in Table 1.

The termination criterion is activated at the second iteration, so that the NP-GBD will have identified $(y, x) = (0.9, 8.52)$ as the "optimum".

3.5. *Convergence to nonextremum points*

Let us also consider $\bar{y} = 1.4$ as a starting point and apply the NP-GBA. The algorithm will converge at the point $(y, x) = (0.92484, 9.0308)$, which is not a local extremum. Details of the iterations are shown in Table 1.

At the first iteration, the master gives a solution with a value above the two local minima. It is therefore impossible to reach any of these minima as iterations progress.

In view of this dissapointing result, words of caution must be raised for the use of the NP-GBA in cases where *Property (P)* does not hold.

3.6. *The use of NP-GBA when L-Dual-Adequacy is not satisfied*

We present now an alternative formulation of our example that satisfies *Property (P)*. By solving the example without using an *L-Dual-Adequacy* algorithm, i.e. an algorithm that does not identify the global minimum in L^*, we show that the same results as before are obtained.

Rewrite the example in the following form:

$$\min_{z_1, z_2} F(z_1, z_2) = f(z_1) - z_2,$$

$$\begin{aligned} \text{s.t.} \quad & z_2 - g(z_1) \leqslant 0, \\ & z_2 + z_1 \leqslant a, \\ & y - z_1 = 0, \\ & y, z_1 \geqslant 0.9, \\ & z_1, z_2 \in X = R, \\ & y \in Y = R \end{aligned}$$

Table 1. Results of the NP-GBA

$\bar{y}$	z_2^*	F_w^*	A^*	B^*	$\hat{y}$	y_0
1.1	13.0533	−5.12354	−5.9947	0.7920	0.9	−5.28194
0.9	8.5200	−5.11029	−4.0411	−1.1880	0.98669	−5.21328
0.98669	10.3672	−5.16576	−5.0310	−0.1366	1.03786	−5.17275
1.03786	11.5421	−5.15986	−5.5225	0.3495	1.01134	−5.16913
1.01134	10.9253	−5.16603	−5.2781	0.1108	0.99882	−5.16742
0.99882	10.6399	−5.16665	−5.1548	−0.0118	1.00503	−5.16673
1.49	24.6545	−4.78162	−4.8926	0.0745	0.9	−4.82556
0.9	8.5200	−5.11029	−4.0411	−1.1880	0.9	−4.82556
1.4	21.6533	−4.81637	−5.75717	0.67200	0.9	−5.15237
0.9	8.5200	−5.11029	−4.04109	−1.18800	0.92263	−5.13716
0.92263	8.9847	−5.13374	−4.31433	−0.88813	0.92482	−5.13569
0.92482	9.0304	−5.13566	−4.34036	−0.85995	0.92484	−5.13568
0.92484	9.0308	−5.13568	−4.34060	−0.85969	0.92484	−5.13568

and use y as complicating variable. The master problem is now:

$$\min_{y_0, y} y_0,$$

$$\text{s.t.} \quad \min_{z_1, z_2 \in X} \{F(z_1, z_2) + u_1^{(k_1)}[z_2 - g(z_1)] + u_2^{(k_1)}(z_2 + z_1 - a) - u_3^{(k_1)} z_1\} + u_3^{(k_1)} y \leqslant y_0, \ \forall k_1,$$

$$\min_{z_1, z_2 \in X} \{\lambda_1^{(k_2)}[z_2 - g(z_1)] + \lambda_2^{(k_2)}(z_2 + z_1 - a) - \lambda_3^{(k_2)} z_1\} + \lambda_3^{(k_2)} y \leqslant 0, \quad \forall k_2,$$

$$y \geqslant 0.9.$$

Property (P) is now satisfied. Since $f(z_1)$ is nonconvex, there is more than one candidate to satisfy *Property* (P'). Using NP-GBA the same results are obtained as before, i.e. starting at a local extremum (minimum or maximum) the termination criterion is immediately met, and starting at some neighborhood of a local minimum, the local minimum is reached. Finally, there are regions where starting points will lead to results that are not extrema. The numerical results iteration-by-iteration are identical to those shown in Table 1.

3.7. *The local character of NP-GBA*

To explain the above results, it is appropriate to reiterate the conceptual nature of Generalized Benders iterations, namely that the master problem is being replaced by a sequence of relaxed problems, where constraints are being added for each new value of the Lagrange multiplier u. To do this search for Lagrange multipliers in a systematic way, the primal problem is solved. Therefore, the primal problem, even if solved globally, can only provide upper-bounds of the solution and be used as a source for new constraints. The success of the global search proposed by the GBD critically depends on the globality of the solutions of the master problem, which in turn, aside from the satisfaction of *Property* (P), also relies on two conditions:

(a) the $\min_{x \in X}$ in the definitions of L^* and L_* should be global;

(b) the solution of the master itself should be global.

If (a) is not satisfied, as happens in the NP-GBA where *Property* (P') is simple assumed, even global solutions of the relaxed master are not guaranteed to be lower bounds of the global optimum [except of course, when *Property* (P) holds and X is convex, as well as $F(x, y)$ and $G(x, y)$ are convex in x]. Therefore, if *Property* (P') does not hold, the NP-GBA is not guaranteed to provide global results. Similarly, if (b) is not satisfied, then again the relaxed master does not provide lower bounds of the global optimum.

Finally, the convergence to nonextremum points is explained by the fact that *Property* (P') is not even locally satisfied for the starting point. Indeed, in the example we have:

$$\min_{z_1, z_2 \in X} \{f(z_1) - z_2 + u_1^*)1[z_2 - g(z_1)] + (z_2 + z_1 - a) - u_3^* z_1\} \neq f(z_1^*) - z_2^* + u_2^* + u_1^*[z_2^* - g(z_1^*)] + u_2^*(z_2^* + z_1^* - a) - u_3^* z_1^*.$$

As a result, constraints are added to the relaxed master problem which define a feasible region that does not contain any local minimum, and thus the algorithm converges to a boundary point of this region, which is not, of course, a local minimum of (1).

3.8. *A criterion for identifying local optima*

The aforementioned examples suggest that when attempting to solve nonconvex problems using NP-GBA one may generate solutions that are not even local extrema. Since it is desirable to use the NP-GBA as a local extrema finder we need a criterion to determine whether the converged values are local extrema or not.

Consider the following:

FACT:
If the NP-GBA is used and the values of the primal and master problems coincide in the first iteration, then the Kuhn–Tucker necessary conditions of problem (1) are satisfied.

Proof—Let $\tilde{y}$ be the fixed point for which the primal is solved in the first iteration. The Kuhn–Tucker conditions of the resulting primal are:

$$\nabla_x F(x^*, \tilde{y}) + u^{*T} \nabla_x G(x^*, \tilde{y}) = 0$$
$$u_i^* G_i(x^*, \tilde{y}) = 0,$$
$$u^* \geqslant 0, \qquad (9)$$

where (x^*, u^*) are all pairs at which the primal's solution is assumed. Since the solutions of the master and primal are identical in the first iteration, it also holds that:

$$\hat{y}_0 = F(x^*, \tilde{y}),$$
$$\hat{y} = \tilde{y},$$

where $(\hat{y}, \hat{y}_0)$ is the solution of the master. One of the Kuhn–Tucker conditions for the master is then:

$$\nabla_y F(x^*, \hat{y}) + u^{*T} \nabla_y G(x^*, \hat{y}) = 0. \qquad (10)$$

But (9–10) constitute the Kuhn–Tucker conditions for problem (1). O.E.Δ.

The above fact suggests the following:

- CRITERION FOR THE IDENTIFICATION OF LOCAL EXTREMA FOR THE NP-GBA—A given point may be a local extremum of problem (1) only if NP-GBA terminates in one iteration.

In view of this criterion the following step should be added to the NP-GBA:

- *Additional Step in NP-GBA*—Once convergence is achieved, reset all constraint counters K^f and K^i to zero and start the algorithm again using the converged point as a starting point. If termination is immediately achieved, a candidate for a local extremum has been identified.

Remark 3.1

The NP-GBA will not converge to a local maximum, unless the starting point itself is a maximum. The reason for this is that near a local maximum, but not at it, the master problem will identify new points that are in the descent direction, rather than in the ascent one. Take for example the behavior of the NP-GBA when starting at point $\bar{y} = 1.49$, which is close to a local maximum. It converges immediately to the other local maximum, going in the descent direction. If NP-GBA is applied again, at a perturbed value near $y = 0.9$ a different solution with lower objective function is obtained. This suggests that the points identified as candidates of an extremum can be subject to an additional test, which include the application of NP-GBA with a starting point in a small neighborhood of the converged point. A convergence to the same point would indicate that the point in question is a candidate to be a local minimum. If convergence to a different point occurs, the original starting point is a local maximum.

4. THE PL-GENERALIZED BENDERS ALGORITHM

In this section we present a general algorithm, that satisfies all the required properties of the GBD, i.e. it satisfies *Property (P)* and is *L-Dual-Adequate*. We call it *PL-Generalized Benders Algorithm* (PL-GBA). We show that if implemented successfully, i.e. if the different optimization subproblems can be globally solved (which may not always be possible) and upon convergence, it will either provide the global optimum, or upper and lower bounds of the global optimum [if a gap exists between $v(y)$ and its dual]. We also show how a recursive application of this procedure, i.e. the use of converged points as new starting points, may be used to conduct a systematic search in the direction of the optimum.

Let us rewrite the original problem (1) by introducing a variable y_1 in the following way:

$$\begin{aligned} \min_{x, y_1, y} \quad & F(x, y_1), \\ \text{s.t.} \quad & G(x, y_1) \leq 0, \\ & y - y_1 = 0, \\ & x \in X, \\ & y, y_1 \in Y. \end{aligned} \tag{11}$$

If y is selected as the complicating variable, the primal problem is:

$$\begin{aligned} \min_{x, y_1} \quad & F(x, y_1), \\ \text{s.t.} \quad & G(x, y_1) \leq 0, \\ & \bar{y} - y_1 = 0, \\ & x \in X, \\ & y_1 \in Y. \end{aligned} \tag{12}$$

Property (P) is satisfied and the relaxed problem is:

$$\begin{aligned} \min_{y_0, y} \; & y_0 \\ \text{s.t.} \; & \min_{x, y_1 \in X} \{F(x, y_1) + u_1^{(k_1)\mathrm{T}} G(x, y_1) - u_2^{(k_1)\mathrm{T}} y_1\} \\ & \qquad + u_2^{(k_1)\mathrm{T}} y \leq y_0, \quad k_1 = 1, \ldots K^f, \\ & \min_{x, y_1 \in X} \{\lambda_1^{(k_2)\mathrm{T}} G(x, y_1) - \lambda_2^{(k_2)\mathrm{T}} y_1\} \\ & \qquad + \lambda_2^{(k_2)\mathrm{T}} y \leq 0, \quad k_2 = 1, \ldots K^i \\ & y \in Y. \end{aligned} \tag{13}$$

Define problems R and S as follows:

Problem R:

$$\begin{aligned} \min_{x, y_1, R} \quad & R, \\ \text{s.t.} \quad & R = \{F(x, y_1) + u_1^{(k_1)\mathrm{T}} G(x, y_1) - u_2^{(k_1)\mathrm{T}} y_1\} \\ & x \in X, \\ & y_1 \in Y. \end{aligned} \tag{14}$$

Problem S:

$$\begin{aligned} \min_{x, y_1, S} \quad & S, \\ \text{s.t.} \quad & S = \{\lambda_1^{(k_2)\mathrm{T}} G(x, y_1) - \lambda_2^{(k_2)\mathrm{T}} y_1\}, \\ & x \in X, \\ & y_1 \in Y. \end{aligned} \tag{15}$$

When the primal problem is feasible, one solves problem R, whereas if the primal problem is infeasible one solves problem S, updating the corresponding counter (K^f or K^i). Let us call the solutions of these problems $(\hat{x}^{(k_1)}, \hat{y}_1^{(k_1)})$ and $(\hat{x}^{(k_2)}, \hat{y}_1^{(k_2)})$, respectively and denote the optimum values by $\hat{R}^{(k_1)}$ and $\hat{S}^{(k_2)}$, respectively. Therefore we write the master problem as follows:

$$\begin{aligned} \min_{y_0, y} \quad & y_0 \\ \text{s.t.} \quad & \hat{R}^{(k_1)} + u_2^{(k_1)\mathrm{T}} y \leq y_0, \quad k_1 = 1, \ldots K^f, \\ & \hat{S}^{(k_2)} + \lambda_2^{(k_2)\mathrm{T}} y \leq 0, \quad k_2 = 1, \ldots K^i. \end{aligned} \tag{16}$$

If problem (11) does not have a gap between $v(y)$ and its dual, then the PL-GBD algorithm will have ϵ-convergence properties. However, if a dual gap is present, the algorithm may provide consecutive relaxed master's solutions that converge to a solution that is lower than the lowest upper bound provided by the primal. In this implementation of the GBD, the master's solution will always stay below the global optimum and the primal's solutions will always stay above. Therefore for termination criterion given above (Section 3, Floudas *et al.*, 1989) is not needed. The convergence criterion should, however, be modified as follows:

—Let $\hat{y}_0(i-1)$ be the value of the relaxed master in the $(i-1)$th iteration. If $U \leqslant \hat{y}_0(i)+\epsilon$, or if $\hat{y}_0(i) \leqslant \hat{y}_0(i-1)+\epsilon$, terminate.

The above procedure not only satisfies *Property (P)* but is also *L-Dual-Adequate* as long as problems R and S are solved for their global solutions. Under these conditions the master (which is a linear programming problem) provides always a valid lower bound of the global optimum (weak duality theorem). Numerical implementation of this procedure requires that the employed values of $\hat{R}^{(k_1)}$ and $\hat{S}^{(k_2)}$ represent global optima, something that is not always possible to guarantee.

Remark 4.1

It may seem that problems R and S are as hard to solve as the original problem. This is not necessarily true. Consider the quadratic problem:

$$\begin{aligned} \min_{x,y} \quad & x^{\mathrm{T}} Q_0 y, \\ \text{s.t.} \quad & x^{T} Q_1 y \leqslant 0, \\ & x \in X, \\ & y \in Y, \end{aligned}$$

with X and Y being closed bounded sets defined by linear constraints and Q_0, Q_1 indefinite matrices.

If y is considered as the complicating variable and the PL algorithm is implemented, problem R is an indefinite quadratic minimization problem subject to linear constraints, whereas the original problem is a quadratic problem with quadratic constraints. While the latter is difficult to solve globally there are algorithms in the literature for the former (Ritter, 1966).

4.1. *Obtaining dual-gaps in nonconvex problems*

As explained above, when a gap is present between $v(y)$ and its dual, the solutions of the primal and the master do not approach one another. Thus, at the conclusion of the Benders iterations, the former is greater than the latter.

Applied to our example with starting point $\bar{y}=1.1$ the procedure provides the results given in Table 2. Since $x^* < a - y$ for all the iterations we obtain $u_1^* = 1$ and $u_2^* = 0$ in all cases. Therefore the variable x cancels out formally in the expression of R, and $\hat{x}$ need not be calculated.

In the last iteration, the master repeats its previous value. Therefore, unless new criteria are added to select new values of u^*, the iterations cannot progress. A gap in the Benders iterations has been identified: the value of the objective function at the global optimum is in the interval $(-5.123, -5.548)$. The identified value is $(y, x) = (1.268, 17.614)$, which is not an extremum of the problem.

Table 2. Results of the PL-GBA

$\bar{y}$	$F(\bar{y}, y_1^*)$	y	y_0
1.100	−5.123	0.900	−6.219
0.900	−5.110	1.460	−5.755
1.460	−4.787	1.268	−5.548
1.268	−4.943	1.268	−5.548
1.400	−4.8163	0.900	−6.0830
0.900	−5.1103	1.423	−5.7316
1.423	−4.8024	1.375	−5.6745
1.375	−4.8349	1.375	−5.6745
1.9	−5.2549	0.900	−49.1536
0.9	−5.1103	1.87686	−6.2708
1.87686	−4.2214	1.87686	−6.2708

Remark 4.2

Problem R does not show unbounded minima for this example. This is not always true.

Remark 4.3

Problem R has more than one local minimum. The unconstrained problem can be solved globally for this example. This, as said, is not always easily achieved.

We show next the results obtained when the starting point is $\bar{y}=1.4$. The iterations are shown in Table 2.

A gap in the Benders iterations has been identified again. The global solution has an objective function value in the interval $(-5.1103, -5.6745)$. The identified point of convergence is $(y, x) = (1.375, 20.851)$ which is the point at which the supporting hyperplane of the dual of $v(1.375)$ touches the feasible region. Similar results, showing the presence of this gap are obtained for other starting points. Even when the starting point is the global optimum, $\bar{y}=1.9$, a gap is identified and $y = 1.87686$ is obtained as the solution. Details of these iterations, shown in Table 2, suggest that the size of this gap depends on the starting point.

4.2. *Recursive application of PL-GBA*

Let y^+ be the value obtained when PL-GBA has concluded. Suppose that the PL-GBA is applied again using y^+ as starting point. This was done for our example using a value of $\epsilon = 1 \times 10^{-7}$, and the results are shown in Table 3. The number of iterations required for each step are indicated in the last column.

Table 3. Results of the recursive application of the PL-GBA

$\bar{y}$	U	y^+	y_0^+	Iter
1.400	−5.1103	1.375	−5.6745	4
1.375	−5.1103	1.253	−5.5305	4
1.253	−5.1103	1.861	−5.5261	4
1.861	−5.2444	1.892	−6.5252	15
1.892	−5.2452	1.893	−6.5220	13
1.893	−5.2452	1.893	−6.5219	11

Several other starting points were used for this recursive application of the PL-GBA and the same result was always obtained. The algorithm converges to $y = 1.893$ with a global optimum in the interval $(-6.5219, -5.2452)$. Note however, that in the case of $y = 1.4$ used as a starting point better lower bounds are obtained for other converged values y^+. If we take the best value obtained, the global optimum can be identified to be in the interval $(-5.5261, -5.2452)$. Depending on the starting point different bounding intervals can be identified using this procedure.

4.3. *Analysis of the results*

One important remark is that the PL-GBA requires the global solution problems R and S. Thus, the PL-GBA cannot always be implemented. If, however, the PL-GBA is implemented successfully, it is capable of identifying bounds on the global optima for nonconvex problems with dual gaps. It is apparent that the use of PL-GBA in the recursive form shown above can complement the results of the NP-GBA, which is a local solver. The solutions of the former may be good starting points of NP-GBA, for the identification of global optima.

5. SUMMARY

The conceptual steps undertaken in establishing the various implementations of GBD are summarized as follows:

—Convexity of X and of $F(x, y)$ and $G(x, y)$ in x, implies that (1) is equivalent to (4). When these conditions do not apply a gap between $v(y)$ and its dual may exist (Remark 2.1).

—Problem (4) is replaced by a sequence of relaxed master problems (Remark 2.3).

—Convexity of X and of $F(x, y)$ and $G(x, y)$ in x (as well as satisfaction of certain other conditions) guarantees ϵ-convergence of the GBD iterations (Geoffrion, 1972, Theorem 2.5). It is implicitly understood, that for these results to hold, both the primal and the relaxed master problem must be solved globally.

—Global solution of the primal stems readily from convexity of X and of $F(x, y)$ and $G(x, y)$ in x. Global solution of the relaxed master may come either through the use of special algorithms or by the establishment of convexity of $L^*(y, u)$ and $L_*(y, \lambda)$ in y. One case for which convexity of these functions can be established is when $F(x, y)$ and $G(x, y)$ are jointly convex in x and y (Geoffrion, 1972, Section 4.2).

—Prior to the solution of the relaxed master problem, the evaluation of $L^*(y, u^*)$ and $L_*(y, \lambda^*)$ is first required through (5) and (6).

—One case for which $L^*(y, u^*)$ and $L_*(y, \lambda^*)$ can be obtained in explicit form is when the global minimum (over x) of (5) and (6) can be obtained independently of y [*Property* (*P*)]. In this case one can construct the Benders iterations by selecting x to be the solution of the primal and u (or λ) the corresponding Lagrange multiplier.

—When the Benders iterations are built based on the primal solution, i.e. assuming that the solution of the minimization problem (5) or (6) is equal to the solution of the primal x^*, [irrespective of the satisfaction of *Property* (*P*)], the NP-GBD algorithm is obtained. When x^* is indeed the solution of the minimization problem, it is said that *Property* (*P'*) holds.

—When the Benders iterations are built through the introduction of additional constraints as in (11), and through global solutions of (5) and (6) (problems R and S), then the PL-GBD is obtained.

6. CONCLUSIONS

This review of the Generalized Benders Decomposition (GBD) reveals that some of its implementations (NP-GBD) may only identify local optima, or points that are not even extrema, suggesting that the procedure should be used with caution, especially when global optimality is being sought. More specifically, it is demonstrated that the implementation of the procedure guarantees globality when certain properties are satisfied (primal convexity, global solution of the master, *Property* (*P*) and *L-Dual-Adequacy*). For the cases in which the *GBD* is implemented without satisfying *Property* (*P*) and/or *L-Dual-Adequacy*, a criterion for determining if the converged value is a local extremum, is provided.

An alternative approach for nonconvex problems is also presented (PL-GBD). The algorithm satisfies *Property* (*P*) and is *L-Dual-Adequate*, but is not always implementable. It has also been shown that the presence of gaps between the projection $v(y)$ and its dual, may prevent this implementation of the GBD from converging to the global solution, providing instead only bounds on the global optimum.

Acknowledgements—Financial support from the UCLA-NSF-ERC under Grant No. CDR 86-22184 and the NSF-PYI under Grant No. CBT 8857867 are gratefully acknowledged.

REFERENCES

Benders J. F., Partitioning procedures for solving mixed variables programming problems. *Numer. Math.* **4**, 238 (1962).

Floudas C. A., A. Aggarwal and A. R. Ciric, Global optimum search for nonconvex NLP and MINLP problems. *Computers chem. Engng* **13**, 10 (1989).

Geoffrion A. M., Duality in nonlinear programming: a simplified application-oriented development. *SIAM Rev.* **13**, 1(1971).

Geoffrion A. M., Generalized Benders Decomposition. *J. Optimization Theory Applic.* **10**, 237 (1972).

Luenberger D. G., *Optimization by Vector Space Methods*. Wiley, New York (1969).

Ritter R., A method for solving maximum problems with a nonconcave quadratic objective function. *Z. Wahrscheinlichkeitstherie. Verw. Geb.* **4**, 340–351 (1966).
Rockafellar R. T., *Convex Analysis*. Princeton University Press, Princeton, NJ (1972).

APPENDIX

The purpose of this Appendix is to rederive Theorem 2.2 of Geoffrion (1972), under the conditions that $G(x, y)$ is not convex with respect to x and/or X is not a convex set.

Theorem

Let $W(y)$ be defined by $W(y) = \{(w_1, w_2): w_1 = G(x, y), w_2 = F(x, y); x \in X\}$. Assume X is a nonempty convex set. Assume further that the set $\Gamma_y = \{\gamma \in R^m: G(x, y) \leqslant \gamma$ for some $x \in X\}$ is closed for each fixed $y \in Y$. Then, a point $\bar{y} \in Y$. Then, a point $\bar{y} \in Y$ is also in the set V if and only if $\bar{y}$ satisfies:

$$\min_{x \in X} \lambda^T G(x, y) \leqslant 0, \quad \forall \lambda \in \Lambda. \tag{A1}$$

Proof—Let $\bar{y} \in V$. Verification of (A1) is straightforward. The converse is proven as follows. Suppose $\bar{y}$ satisfies (A1). Then:

$$\max_{\lambda \in \Lambda} \min_{x \in X} [\lambda^T G(x, \bar{y})] \leqslant 0. \tag{A2}$$

By rescaling λ appropriately one obtains:

$$\max_{\lambda > 0} \min_{x \in X} [\lambda^T G(x, \bar{y})\} \leqslant 0. \tag{A3}$$

However, (A3) is the dual of the following problem:

$$\min_{x \in X} \ 0^T x,$$
$$\text{s.t.} \quad G(x, \bar{y}) \leqslant 0. \tag{A4}$$

Suppose that the region $W(y)$ satisfies the inequality $w_1 \geqslant a \geqslant 0$. Then, $w_1 = a$ is a supporting hyperplane, which intercepts the axis $w_1 = 0$ at infinity. Therefore the dual problem would have $+\infty$ as a solution. Since the dual has a finite optimum, and Γ_y is closed, we conclude that $a \leqslant 0$. Therefore there exists at least one point for which the primal problem is feasible.