Chapter 6 Mixed-Integer Nonlinear Optimization

This chapter presents the fundamentals and algorithms for mixed-integer nonlinear optimization problems. Sections 6.1 and 6.2 outline the motivation, formulation, and algorithmic approaches. Section 6.3 discusses the Generalized Benders Decomposition and its variants. Sections 6.4, 6.5 and 6.6 presents the Outer Approximation and its variants with Equality Relaxation and Augmented Penalty. Section 6.7 discusses the Generalized Outer Approximation while section 6.8 compares the Generalized Benders Decomposition with the Outer Approximation. Finally, section 6.9 discusses the Generalized Cross Decomposition.

6.1 Motivation

A wide range of nonlinear optimization problems involve integer or discrete variables in addition to the continuous variables. These classes of optimization problems arise from a variety of applications and are denoted as Mixed-Integer Nonlinear Programming MINLP problems.

The integer variables can be used to model, for instance, sequences of events, alternative candidates, existence or nonexistence of units (in their zero-one representation), while discrete variables can model, for instance, different equipment sizes. The continuous variables are used to model the input-output and interaction relationships among individual units/operations and different interconnected systems.

The nonlinear nature of these mixed–integer optimization problems may arise from (i) nonlinear relations in the integer domain exclusively (e.g., products of binary variables in the quadratic assignment model), (ii) nonlinear relations in the continuous domain only (e.g., complex nonlinear input-output model in a distillation column or reactor unit), (iii) nonlinear relations in the joint integer-continuous domain (e.g., products of continuous and binary variables in the scheduling/planning of batch processes, and retrofit of heat recovery systems). In this chapter, we will focus on nonlinearities due to relations (ii) and (iii). An excellent book that studies mixed-integer linear optimization, and nonlinear integer relationships in combinatorial optimization is the one by Nemhauser and Wolsey (1988).

The coupling of the integer domain with the continuous domain along with their associated nonlinearities make the class of MINLP problems very challenging from the theoretical, algorithmic, and computational point of view. Apart from this challenge, however, there exists a broad spectrum of applications that can be modeled as mixed-integer nonlinear programming problems. These applications have a prominent role in the area of *Process Synthesis* in chemical engineering and include: (i) the synthesis of grassroot heat recovery networks (Floudas and Ciric, 1989; Ciric and Floudas, 1990; Ciric and Floudas, 1989; Yee *et al.*, 1990a; Yee and Grossmann, 1990; Yee *et al.*, 1990b); (ii) the retrofit of heat exchanger systems (Ciric and Floudas, 1990; Papalexandri and Pistikopoulos, 1993); (iii) the synthesis of distillation-based separation systems (Paules and Floudas, 1988; Viswanathan and Grossmann, 1990; Aggarwal and Floudas, 1990; Aggarwal and Floudas, 1992; Paules and Floudas, 1992); (iv) the synthesis of complex reactor networks (Kokossis and Floudas, 1990; Kokossis and Floudas, 1994); (v) the synthesis of reactor-separator-recycle systems (Kokossis and Floudas, 1991); (vi) the synthesis of utility systems (Kalitventzeff and Marechal, 1988); and the synthesis of total process systems (Kocis and Grossmann, 1988; Kocis and Grossmann, 1989a; Kravanja and Grossmann, 1990). An excellent review of the mixed–integer nonlinear optimization frameworks and applications in *Process Synthesis* are provided in Grossmann (1990). Algorithmic advances for logic and global optimization in Process Synthesis are reviewed in Floudas and Grossmann (1994).

Key applications of MINLP approaches have also emerged in the area of *Design, Scheduling, and Planning of Batch Processes* in chemical engineering and include: (i) the design of multiproduct plants (Grossmann and Sargent, 1979; Birewar and Grossmann, 1989; Birewar and Grossmann, 1990); and (ii) the design and scheduling of multipurpose plants (Vaselenak *et al.*, 1987; Vaselenak *et al.*, 1987; Faqir and Karimi, 1990; Papageorgaki and Reklaitis, 1990; Papageorgaki and Reklaitis, 1990; Wellons and Reklaitis, 1991; Wellons and Reklaitis, 1991; Sahinidis and Grossmann, 1991; Fletcher *et al.*, 1991). Excellent recent reviews of the advances in the design, scheduling, and planning of batch plants can be found in Reklaitis (1991), and Grossmann *et al.* (1992).

Another important applications of MINLP models have recently been reported for (i) the computer-aided molecular design aspects of selecting the best solvents (Odele and Macchietto, 1993); and (ii) the interaction of design and control of chemical processes (Luyben and Floudas (1994a), Luyben and Floudas (1994b)).

MINLP applications received significant attention in other engineering disciplines. These include (i) the facility location in a multiattribute space (Ganish *et al.*, 1983); (ii) the optimal unit allocation in an electric power system (Bertsekas *et al.*, 1983); (iii) the facility planning of an electric power generation (Bloom, 1983; Rouhani *et al.*, 1985); (iv) the topology optimization of transportation networks (Hoang, 1982); and (v) the optimal scheduling of thermal generating units (Geromel and Belloni, 1986).

6.2 Formulation

The primary objective in this section is to present the general formulation of MINLP problems, discuss the difficulties and present an overview of the algorithmic approaches developed for the solution of such models.

6.2.1 Mathematical Description

The general MINLP formulation can be stated as

$$\begin{aligned} \min_{x,y} \quad & f(x,y) \\ s.t. \quad h(x,y) &= 0 \\ g(x,y) &\leq 0 \\ x &\in X \subseteq \Re^n \\ y &\in Y \text{ integer} \end{aligned} \tag{6.1}$$

Here x represents a vector of n continuous variables (e.g., flows, pressures, compositions, temperatures, sizes of units), and y is a vector of integer variables (e.g., alternative solvents or materials); $h(x,y) = \mathbf{0}$ denote the m equality constraints (e.g., mass, energy balances, equilibrium relationships); $g(x,y) \leq \mathbf{0}$ are the p inequality constraints (e.g., specifications on purity of distillation products, environmental regulations, feasibility constraints in heat recovery systems, logical constraints); $f(x,y)$ is the objective function (e.g., annualized total cost, profit, thermodynamic criteria).

Remark 1 The integer variables y with given lower and upper bounds,

$$y^L \leq y \leq y^U,$$

can be expressed through 0–1 variables (i.e., binary) denoted as $\mathbf{z}$, by the following formula:

$$y = y^L + z_1 + 2z_2 + 4z_3 + \ldots + 2^{N-1}z_N,$$

where N is the minimum number of 0–1 variables needed. This minimum number is given by

$$N = 1 + INT\left\{\frac{\log(y^U - y^L)}{\log 2}\right\},$$

where the INT function truncates its real argument to an integer value. This approach however may not be practical when the bounds are large.

Then, formulation (6.1) can be written in terms of 0–1 variables:

$$\begin{aligned} \min_{x,y} \quad & f(x,y) \\ \text{s.t.} \quad h(x,y) &= 0 \\ g(x,y) &\leq 0 \\ x &\in X \subseteq \Re^n \\ y &\in Y = \{0,1\}^q \end{aligned} \tag{6.2}$$

where y now is a vector of q 0–1 variables (e.g., existence of a process unit ($y_i = 1$) or non-existence ($y_i = 0$). We will focus on (6.2) in the majority of the subsequent developments.

6.2.2 Challenges/Difficulties in MINLP

Dealing with mixed–integer nonlinear optimization models of the form (6.1) or (6.2) present two major challenges/difficulties. These difficulties are associated with the nature of the problem, namely, the combinatorial domain (y–domain) and the continuous domain (x–domain).

As the number of binary variables y in (6.2) increase, one faces with a large combinatorial problem, and the complexity analysis results characterize the MINLP problems as NP-complete (Nemhauser and Wolsey, 1988). At the same time, due to the nonlinearities the MINLP problems are in general nonconvex which implies the potential existence of multiple local solutions. The determination of a global solution of the nonconvex MINLP problems is also NP-hard (Murty and Kabadi, 1987), since even the global optimization of constrained nonlinear programming problems can be NP-hard (Pardalos and Schnitger, 1988), and even quadratic problems with one negative eigenvalue are NP-hard (Pardalos and Vavasis, 1991). An excellent book on complexity issues for nonlinear optimization is the one by Vavasis (1991).

Despite the aforementioned discouraging results from complexity analysis which are worst-case results, significant progress has been achieved in the MINLP area from the theoretical, algorithmic, and computational perspective. As a result, several algorithms have been proposed, their convergence properties have been investigated, and a large number of applications now exist that cross the boundaries of several disciplines. In the sequel, we will discuss these developments.

6.2.3 Overview of MINLP Algorithms

A representative collection of algorithms developed for solving MINLP models of the form (6.2) or restricted classes of (6.2) includes, in chronological order of development, the following:

1. Generalized Benders Decomposition, **GBD** (Geoffrion, 1972; Paules and Floudas, 1989; Floudas *et al.*, 1989);
2. Branch and Bound, **BB** (Beale, 1977; Gupta, 1980; Ostrovsky *et al.*, 1990; Borchers and Mitchell, 1991);
3. Outer Approximation, **OA** (Duran and Grossmann, 1986a);
4. Feasibility Approach, **FA** (Mawengkang and Murtagh, 1986);
5. Outer Approximation with Equality Relaxation, **OA/ER** (Kocis and Grossmann, 1987);
6. Outer Approximation with Equality Relaxation and Augmented Penalty, **OA/ER/AP** (Viswanathan and Grossmann, 1990)
7. Generalized Outer Approximation, **GOA** (Fletcher and Leyffer, 1994)
8. Generalized Cross Decomposition, **GCD** (Holmberg, 1990);

In the pioneering work of Geoffrion (1972) on the Generalized Benders Decomposition **GBD** two sequences of updated upper (nonincreasing) and lower (nondecreasing) bounds are created that converge within ϵ in a finite number of iterations. The upper bounds correspond to solving subproblems in the x variables by fixing the y variables, while the lower bounds are based on duality theory.

The branch and bound **BB** approaches start by solving the continuous relaxation of the MINLP and subsequently perform an implicit enumeration where a subset of the 0–1 variables is fixed at each node. The lower bound corresponds to the NLP solution at each node and it is used to expand on the node with the lowest lower bound (i.e., breadth first enumeration), or it is used to eliminate nodes if the lower bound exceeds the current upper bound (i.e., depth first enumeration). If the continuous relaxation NLP of the MINLP has 0–1 solution for the y variables, then the **BB** algorithm will terminate at that node. With a similar argument, if a tight NLP relaxation results in the first node of the tree, then the number of nodes that would need to be eliminated can be low. However, loose NLP relaxations may result in having a large number of NLP subproblems to be solved which do not have the attractive update features that LP problems exhibit.

The Outer Approximation **OA** addresses problems with nonlinear inequalities, and creates sequences of upper and lower bounds as the **GBD**, but it has the distinct feature of using primal information, that is the solution of the upper bound problems, so as to linearize the objective and constraints around that point. The lower bounds in **OA** are based upon the accumulation of the linearized objective function and constraints, around the generated primal solution points.

The feasibility approach **FA** rounds the relaxed NLP solution to an integer solution with the least local degradation by successively forcing the superbasic variables to become nonbasic based on the reduced cost information.

The **OA/ER** algorithm, extends the **OA** to handle nonlinear equality constraints by relaxing them into inequalities according to the sign of their associated multipliers.

The **OA/ER/AP** algorithm introduces an augmented penalty function in the lower bound subproblems of the **OA/ER** approach.

The Generalized Outer Approximation **GOA** extends the **OA** to the MINLP problems of types (6.1),(6.2) and introduces exact penalty functions.

The Generalized Cross Decomposition **GCD** simultaneously utilizes primal and dual information by exploiting the advantages of Dantzig-Wolfe and Generalized Benders decomposition.

In the subsequent sections, we will concentrate on the algorithms that are based on decomposition and outer approximation, that is on 1., 3., 5., 6., 7., and 8.. This focus of our study results from the existing evidence of excellent performance of the aforementioned decomposition-based and outer approximation algorithms compared to the branch and bound methods and the feasibility approach.

6.3 Generalized Benders Decomposition, GBD

6.3.1 Formulation

Geoffrion (1972) generalized the approach proposed by Benders (1962), for exploiting the structure of mathematical programming problems (6.2), to the class of optimization problems stated as

$$
\begin{aligned}
\min_{x,y} \quad & f(x,y) \\
s.t. \quad & h(x,y) = 0 \\
& g(x,y) \leq 0 \\
& x \in X \subseteq \Re^n \\
& y \in Y = \{0,1\}^q
\end{aligned}
$$

under the following conditions:

C1: X is a nonempty, convex set and the functions

$$
\begin{aligned}
f : \Re^n \times \Re^q &\rightarrow \Re, \\
g : \Re^n \times \Re^q &\rightarrow \Re^p,
\end{aligned}
$$

are convex for each fixed $y \in Y = \{0,1\}^q$, while the functions $h : \Re^n \times \Re^l \rightarrow \Re^m$ are linear for each fixed $y \in Y = \{0,1\}^q$.

C2: The set

$$Z_y = \{z \in \Re^p : h(x,y) = 0,\ g(x,y) \leq z \ \text{ for some } \ x \in X\},$$

is closed for each fixed $y \in Y$.

C3: For each fixed $y \in Y \cap V$, where

$$V = \{y : h(x,y) = 0,\ g(x,y) \leq 0,\ \text{for some } \ x \in X\},$$

one of the following two conditions holds:

(i) the resulting problem (6.2) has a finite solution and has an optimal multiplier vector for the equalities and inequalities.

(ii) the resulting problem (6.2) is unbounded, that is, its objective function value goes to $-\infty$.

Remark 1 It should be noted that the above stated formulation (6.2) is, in fact, a subclass of the problems for which the **GBD** of Geoffrion (1972) can be applied. This is due to the specification of $y = \{0,1\}$, while Geoffrion (1972) investigated the more general case of $Y \subseteq \Re^q$, and defined the vector of y variables as "complicating" variables in the sense that if we fix y, then:

(a) Problem (6.2) may be decomposed into a number of independent problems, each involving a different subvector of $\boldsymbol{x}$; or

(b) Problem (6.2) takes a well known special structure for which efficient algorithms are available; or

(c) Problem (6.2) becomes convex in $\boldsymbol{x}$ even though it is nonconvex in the joint $\boldsymbol{x}$–$\boldsymbol{y}$ domain; that is, it creates special structure.

Case (a) may lead to parallel computations of the independent subproblems. Case (b) allows the use of special-purpose algorithms (e.g., generalized network algorithms), while case (c) invokes special structure from the convexity point of view that can be useful for the decomposition of non-convex optimization problems Floudas *et al.* (1989).

In the sequel, we concentrate on $\boldsymbol{Y} = \{0, 1\}^q$ due to our interest in MINLP models. Note also that the analysis includes the equality constraints $\boldsymbol{h}(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{0}$ which are not treated explicitly in Geoffrion (1972).

Remark 2 Condition C2 is not stringent, and it is satisfied if one of the following holds (in addition to C1,C3):

(i) $\boldsymbol{x}$ is bounded and closed and $\boldsymbol{h}(\boldsymbol{x}, \boldsymbol{y}), \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{y})$ are continuous on $\boldsymbol{x}$ for each fixed $\boldsymbol{y} \in \boldsymbol{Y}$.

(ii) There exists a point $z_{\boldsymbol{y}}$ such that the set

$$\{\boldsymbol{x} \in \boldsymbol{X} : \; \boldsymbol{h}(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{0}, \; \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{y}) \leq z_{\boldsymbol{y}}\}$$

is bounded and nonempty.

Note though that mere continuity of $\boldsymbol{h}(\boldsymbol{x}, \boldsymbol{y}),\; \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{y})$ on $\boldsymbol{X}$ for each fixed $\boldsymbol{y} \in \boldsymbol{Y}$ does not imply that condition C2 is satisfied. For instance, if $\boldsymbol{X} = [1, \infty]$ and $h(x, y) = x + y$, $g(x, y) = -\frac{1}{x}$, then $z_y = (-\infty, 0)$ which is not closed since for $x \to \infty, g(x, y) \to -\infty$.

Remark 3 Note that the set $\boldsymbol{V}$ represents the values of $\boldsymbol{y}$ for which the resulting problem (6.2) is feasible with respect to $\boldsymbol{x}$. In others words, $\boldsymbol{V}$ denotes the values of $\boldsymbol{y}$ for which there exists a feasible $\boldsymbol{x} \in \boldsymbol{X}$ for $\boldsymbol{h}(\boldsymbol{x}, \boldsymbol{y}) = 0, \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{y}) \leq 0$. Then, the intersection of $\boldsymbol{y}$ and $\boldsymbol{V}$, $\boldsymbol{Y} \cap \boldsymbol{V}$, represents the projection of the feasible region of (2) onto the $\boldsymbol{y}$–space.

Remark 4 Condition C3 is satisfied if a first-order constraint qualification holds for the resulting problem (6.2) after fixing $\boldsymbol{y} \in \boldsymbol{Y} \cap \boldsymbol{V}$.

6.3.2 Basic Idea

The basic idea in Generalized Benders Decomposition **GBD** is the generation, at each iteration, of an upper bound and a lower bound on the sought solution of the MINLP model. The upper bound results from the *primal* problem, while the lower bound results from the *master* problem. The primal problem corresponds to problem (6.2) with fixed $\boldsymbol{y}$–variables (i.e., it is in the $\boldsymbol{x}$–space only), and its solution provides information about the upper bound and the Lagrange multipliers associated with the equality and inequality constraints. The master problem is derived via nonlinear duality theory, makes use of the Lagrange multipliers obtained in the primal problem, and its solution provides information about the lower bound, as well as the next set of fixed y–variables to be used subsequently in the primal problem. As the iterations proceed, it is shown that the sequence of updated upper bounds is nonincreasing, the sequence of lower bounds is nondecreasing, and that the sequences converge in a finite number of iterations.

6.3.3 Theoretical Development

This section presents the theoretical development of the Generalized Benders Decomposition **GBD**. The primal problem is analyzed first for the feasible and infeasible cases. Subsequently, the theoretical analysis for the derivation of the master problem is presented.

6.3.3.1 The Primal Problem

The primal problem results from fixing the y variables to a particular 0–1 combination, which we denote as y^k where k stands for the iteration counter. The formulation of the primal problem $P(y^k)$, at iteration k is

$$\left[\begin{array}{lrcl} \min\limits_{x} & f(x,y^k) & & \\ s.t. & h(x,y^k) & = & 0 \\ & g(x,y^k) & \leq & 0 \\ & x \in X & \subseteq & \Re^n \end{array}\right. \qquad (P(y^k))$$

Remark 1 Note that due to conditions C1 and C3(i), the solution of the primal problem $P(y^k)$ is its global solution.

We will distinguish the two cases of (i) feasible primal, and (ii) infeasible primal, and describe the analysis for each case separately.

Case (i): Feasible Primal
If the primal problem at iteration k is feasible, then its solution provides information on x^k, $f(x^k, y^k)$, which is the upper bound, and the optimal multiplier vectors λ^k, μ^k for the equality and inequality constraints. Subsequently, using this information we can formulate the Lagrange function as

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

Case (ii): Infeasible Primal
If the primal is detected by the NLP solver to be infeasible, then we consider its constraints

$$\begin{array}{rcl} h(x,y^k) & = & 0 \\ g(x,y^k) & \leq & 0 \\ x \in X & \subseteq & \Re^n \end{array}$$

where the set $\boldsymbol{X}$, for instance, consists of lower and upper bounds on the $\boldsymbol{x}$ variables. To identify a feasible point we can minimize an l_1 or l_∞ sum of constraint violations. An l_1-minimization problem can be formulated as

$$\begin{aligned} \min_{\boldsymbol{x} \in \boldsymbol{X}} \quad & \sum_{i=1}^{p} \alpha_i \\ s.t. \quad \boldsymbol{h}(\boldsymbol{x}, \boldsymbol{y}^k) &= \boldsymbol{0} \\ g_i(\boldsymbol{x}, \boldsymbol{y}^k) &\leq \alpha_i \quad i = 1, 2, \ldots, p \\ \alpha_i &\geq 0 \quad i = 1, 2, \ldots, p \end{aligned}$$

Note that if $\sum_{i=1}^{p} \alpha_i = 0$, then a feasible point has been determined.

Also note that by defining as

$$\begin{aligned} \alpha^+ &= \max(0, \alpha) \quad \text{and} \\ g_i^+(\boldsymbol{x}, \boldsymbol{y}^k) &= \max\left[0, g_i(\boldsymbol{x}, \boldsymbol{y}^k)\right] \end{aligned}$$

the l_1–minimization problem is stated as

$$\begin{aligned} \min_{\boldsymbol{x} \in \boldsymbol{X}} \quad & \sum_{i=1}^{P} g_i^+ \\ s.t. \quad \boldsymbol{h}(\boldsymbol{x}, \boldsymbol{y}^k) &= \boldsymbol{0} \end{aligned}$$

An l_∞–minimization problem can be stated similarly as:

$$\begin{aligned} \min_{\boldsymbol{x} \in \boldsymbol{X}} \max_{1,2,\ldots,p} \quad & g_i^+(\boldsymbol{x}, \boldsymbol{y}^k) \\ s.t. \quad \boldsymbol{h}(\boldsymbol{x}, \boldsymbol{y}^k) &= \boldsymbol{0} \end{aligned}$$

Alternative feasibility minimization approaches aim at keeping feasibility in any constraint residual once it has been established. An l_1-minimization in these approaches takes the form:

$$\begin{aligned} \min_{\boldsymbol{x} \in \boldsymbol{X}} \quad & \sum_{i \in \boldsymbol{I}'} g_i^+(\boldsymbol{x}, \boldsymbol{y}^k) \\ s.t. \quad \boldsymbol{h}(\boldsymbol{x}, \boldsymbol{y}^k) &= \boldsymbol{0} \\ g_i(\boldsymbol{x}, \boldsymbol{y}^k) &\leq 0, \quad i \in \boldsymbol{I} \end{aligned}$$

where $\boldsymbol{I}$ is the set of feasible constraints; and $\boldsymbol{I}'$ is the set of infeasible constraints. Other methods seek feasibility of the constraints one at a time whilst maintaining feasibility for inequalities indexed by $i \in \boldsymbol{I}$. This feasibility problem is formulated as

$$\min_{x \in X} \sum_{i \in I'} w_i \, g_i^+(x, y^k)$$
$$s.t. \quad h(x, y^k) = 0$$
$$g_i(x, y^k) \leq 0, \quad i \in I$$

and it is solved at any one time.

To include all mentioned possibilities Fletcher and Leyffer (1994) formulated a general feasibility problem **(FP)** defined as

$$\left[\begin{array}{ll} \min\limits_{x \in X} & \sum\limits_{i \in I'} w_i \, g_i^+(x, y^k) \\ s.t. & h(x, y^k) = 0 \\ & g_i(x, y^k) \leq 0, \quad i \in I \end{array}\right. \qquad \textbf{(FP)}$$

The weights w_i are non-negative and not all are zero. Note that with $w_i = 1 \;\; i \in I'$ we obtain the l_1–minimization. Also in the l_∞–minimization, there exist nonnegative weights at the solution such that

$$\sum_{i \in I'} w_i = 1,$$

and $w_i = 0$ if $g_i(x, y^k)$ does not attain the maximum value.

Note that infeasibility in the primal problem is detected when a solution of **(FP)** is obtained for which its objective value is greater than zero.

The solution of the feasibility problem **(FP)** provides information on the Lagrange multipliers for the equality and inequality constraints which are denoted as $\bar{\lambda}^k, \bar{\mu}^k$ respectively. Then, the Lagrange function resulting from on infeasible primal problem at iteration k can be defined as

$$\bar{L}^k(x, y, \bar{\lambda}^k, \bar{\mu}^k) = \bar{\lambda}^{k^T} h(x, y) + \bar{\mu}^{k^T} g(x, y).$$

Remark 2 It should be noted that two different types of Lagrange functions are defined depending on whether the primal problem is feasible or infeasible. Also, the upper bound is obtained only from the feasible primal problem.

6.3.3.2 The Master Problem

The derivation of the master problem in the **GBD** makes use of nonlinear duality theory and is characterized by the following three key ideas:

(i) Projection of problem (6.2) onto the y–space;

(ii) Dual representation of V; and

(iii) Dual representation of the projection of problem (6.2) on the y–space.

In the sequel, the theoretical analysis involved in these three key ideas is presented.

(i) Projection of (6.2) onto the y–space

Problem (6.2) can be written as

$$\begin{aligned} \min_{y} \inf_{x} \quad & f(x,y) \\ s.t. \quad h(x,y) &= 0 \\ g(x,y) &\leq 0 \\ x &\in X \\ y &\in Y = \{0,1\}^q \end{aligned} \tag{6.3}$$

where the min operator has been written separately for y and x. Note that it is infimum with respect to x since for given y the inner problem may be unbounded. Let us define $v(y)$ as

$$\begin{aligned} v(y) \;=\; \inf_{x} & \, f(x,y) \\ s.t. \quad h(x,y) &= 0 \\ g(x,y) &\leq 0 \\ x &\in X \end{aligned} \tag{6.4}$$

Remark 1 Note that $v(y)$ is parametric in the y variables and therefore, from its definition corresponds to the optimal value of problem (6.2) for fixed y (i.e., the primal problem $P(y^k)$ for $y = y^k$).

Let us also define the set V as

$$V = \{y : h(x,y) = 0,\; g(x,y) \leq 0 \;\text{ for some }\; x \in X\}\,. \tag{6.5}$$

Then, problem (6.3) can be written as

$$\begin{aligned} \min_{y} \quad & v(y) \\ s.t. \quad & y \in Y \cap V \end{aligned} \tag{6.6}$$

where $v(y)$ and V are defined by (6.4) and (6.5), respectively.

Remark 2 Problem (6.6) is the projection of problem (6.2) onto the y–space. Note also that in (6.5) $y \in Y \cap V$ since the projection needs to satisfy the feasibility considerations.

Having defined the projection problem of (6.2) onto the y–space, we can now state the theoretical result of Geoffrion (1972).

Theorem 6.3.1 (Projection)

(i) *If* (x^*, y^*) *is optimal in (6.2), then* y^* *is optimal in (6.6).*

(ii) *If (6.2) is infeasible or has unbounded solution, then the same is true for (6.6) and vice versa.*

Remark 3 Note that the difficulty in (6.6) is due to the fact that $v(y)$ and V are known only implicitly via (6.4) and (6.5).

To overcome the aforementioned difficulty we have to introduce the dual representation of V and $v(y)$.

(ii) Dual Representation of V

The dual representation of V will be invoked in terms of the intersection of a collection of regions that contain it, and it is described in the following theorem of Geoffrion (1972).

Theorem 6.3.2 (Dual Representation of V)
Assuming conditions C1 and C2 a point $y \in Y$ *belongs also to the set* V *if and only if it satisfies the system:*

$$0 \geq \inf \bar{L}(x, y, \bar{\lambda}, \bar{\mu}), \quad \forall \bar{\lambda}, \bar{\mu} \in \Lambda, \tag{6.7}$$

$$\textit{where} \quad \Lambda = \left\{ \bar{\lambda} \in \Re^m, \bar{\mu} \in \Re^p : \bar{\mu} \geq 0, \sum_{i=1}^{p} \bar{\mu}_i = 1 \right\}.$$

Remark 4 Note that (6.7) is an infinite system because it has to be satisfied for all $\bar{\lambda}, \bar{\mu} \in \Lambda$.

Remark 5 The dual representation of the set V needs to be invoked so as to generate a collection of regions that contain it (i.e., system (6.7) corresponds to the set of constraints that have to be incorporated for the case of infeasible primal problems.

Remark 6 Note that if the primal is infeasible and we make use of the l_1–minimization of the type:

$$\begin{aligned} \min_{x} \quad & \sum_{i \in I} \alpha_i \\ s.t. \quad h(x, y^k) &= 0 \\ g_i(x, y^k) &\leq \alpha_i, \quad i \in I \\ x &\in X \end{aligned} \tag{6.8}$$

then the set Λ results from a straightforward application of the **KKT** gradient conditions to problem (6.8) with respect to α_i.

Having introduced the dual representation of the set V, which corresponds to infeasible primal problems, we can now invoke the dual representation of $v(y)$.

(iii) Dual Representation of $v(y)$

The dual representation of $v(y)$ will be in terms of the pointwise infimum of a collection of functions that support it, and it is described in the following theorem due to Geoffrion (1972).

Theorem 6.3.3 (Dual of $v(y)$)

$$v(y) = \left[\begin{array}{lrcl} \inf_{x} & f(x,y) & & \\ \text{s.t.} & h(x,y) & = & 0 \\ & g(x,y) & \leq & 0 \\ & x & \in & X \end{array} \right] = \left[\sup_{\lambda,\mu \geq 0} \quad \inf_{x \in X} L(x,y,\lambda,\mu) \right], \forall y \in Y \cap V \tag{6.9}$$

$$\textit{where} \quad L(x,y,\lambda,\mu) \;=\; f(x,y) + \lambda^T h(x,y) + \mu^T g(x,y).$$

Remark 7 The equality of $v(y)$ and its dual is due to having the strong duality theorem satisfied because of conditions C1, C2, and C3.

Substituting (6.9) for $v(y)$ and (6.7) for $y \in Y \cap V$ into problem (6.6), which is equivalent to (6.3), we obtain

$$\begin{array}{ll} \min_{y \in Y} & \sup_{\lambda,\mu \geq 0} \inf_{x \in X} L(x,y,\lambda,\mu) \\ s.t. & 0 \geq \inf_{x \in X} \bar{L}(x,y,\bar{\lambda},\bar{\mu}) \end{array}$$

Using the definition of supremum as the lowest upper bound and introducing a scalar μ_B we obtain:

$$\left[\begin{array}{lrll} \min_{y \in Y, \mu_B} & \mu_B & & \\ s.t. & \mu_B & \geq \inf_{x \in X} L(x,y,\lambda,\mu), & \forall \lambda, \forall \mu \geq 0 \\ & 0 & \geq \inf_{x \in X} \bar{L}(x,y,\bar{\lambda},\bar{\mu}), & \forall\, (\bar{\lambda},\bar{\mu}) \in \Lambda \end{array} \right. \tag{M}$$

$$\text{where} \quad \begin{array}{l} L(x,y,\lambda,\mu) = f(x,y) + \lambda^T h(x,y) + \mu^T g(x,y) \\ L(x,y,\bar{\lambda},\bar{\mu}) = \bar{\lambda}^T h(x,y) + \bar{\mu}^T g(x,y) \end{array}$$

which is called the *master* problem and denoted as **(M)**.

Remark 8 If we assume that the optimum solution of $v(y)$ in (6.4) is bounded for all $y \in Y \cap V$, then we can replace the infimum with a minimum. Subsequently, the master problem will be as follows:

$$\begin{aligned} \min_{y \in Y, \mu_B} \quad & \mu_B \\ \text{s.t.} \quad & \mu_B \geq \min_{x \in X} L(x, y, \lambda, \mu), \quad \forall \lambda, \mu \geq 0 \\ & 0 \geq \min_{x \in X} \bar{L}(x, y, \bar{\lambda}, \bar{\mu}), \quad \forall (\bar{\lambda}, \bar{\mu}) \in \Lambda \end{aligned}$$

where $L(x, y, \lambda, \mu)$ and $\bar{L}(x, y, \bar{\lambda}, \bar{\mu})$ are defined as before.

Remark 9 Note that the master problem **(M)** is equivalent to (6.2). It involves, however, an infinite number of constraints, and hence we would need to consider a relaxation of the master (e.g., by dropping a number of constraints) which will represent a lower bound on the original problem. Note also that the master problem features an outer optimization problem with respect to $y \in Y$ and inner optimization problems with respect to x which are in fact parametric in y. It is this outer-inner nature that makes the solution of even a relaxed master problem difficult.

Remark 10 (Geometric Interpretation of the Master Problem) The inner minimization problems

$$\begin{aligned} & \min_{x \in X} L(x, y, \lambda, \mu), \qquad \forall \lambda, \forall \mu \geq 0, \\ & \min_{x \in X} \bar{L}(x, y, \bar{\lambda}, \bar{\mu}), \qquad \forall (\bar{\lambda}, \bar{\mu}) \in \Lambda, \end{aligned}$$

are functions of y and can be interpreted as support functions of $v(y)$ [$\xi(y)$ is a support function of $v(y)$ at point y_o if and only if $\xi(y_o) = v(y_o)$ and $\xi(y) \leq v(y) \; \forall \, y \neq y_o$]. If the support functions are linear in y, then the master problem approximates $v(y)$ by tangent hyperplanes and we can conclude that $v(y)$ is convex in y. Note that $v(y)$ can be convex in y even though problem (6.2) is nonconvex in the joint x–y space Floudas and Visweswaran (1990).

In the sequel, we will define the aforementioned minimization problems in terms of the notion of support functions; that is

$$\begin{aligned} \xi(y; \lambda, \mu) &= \min_{x \in X} L(x, y, \lambda, \mu), \quad \forall \lambda, \; \forall \mu \geq 0 \\ \bar{\xi}(y; \bar{\lambda}, \bar{\mu}) &= \min_{x \in X} \bar{L}(x, y, \bar{\lambda}, \bar{\mu}), \quad \forall (\bar{\lambda}, \bar{\mu}) \in \Lambda \end{aligned}$$

6.3.4 Algorithmic Development

In the previous section we discussed the primal and master problem for the **GBD**. We have the primal problem being a (linear or) nonlinear programming NLP problem that can be solved via available local NLP solvers (e.g., MINOS 5.3). The master problem, however, consists of outer and inner optimization problems, and approaches towards attaining its solution are discussed in the following.

6.3.4.1 How to Solve the Master Problem

The master problem has as constraints the two inner optimization problems (i.e., for the case of feasible primal and infeasible primal problems) which, however, need to be considered for all λ and all $\mu \geq 0$ (i.e., feasible primal) and all $(\bar{\lambda}, \bar{\mu}) \in \Lambda$ (i.e., infeasible). This implies that the master problem has a very large number of constraints.

The most natural approach for solving the master problem is *relaxation* (Geoffrion, 1972). The basic idea in the relaxation approach consists of the following: (i) ignore all but a few of the constraints that correspond to the inner optimization problems (e.g., consider the inner optimization problems for specific or fixed multipliers (λ^1, μ^1) or $(\bar{\lambda}^1, \bar{\mu}^1)$); (ii) solve the relaxed master problem and check whether the resulting solution satisfies all of the ignored constraints. If not, then generate and add to the relaxed master problem one or more of the violated constraints and solve the new relaxed master problem again; (iii) continue until a relaxed master problem satisfies all of the ignored constraints, which implies that an optimal solution at the master problem has been obtained or until a termination criterion indicates that a solution of acceptable accuracy has been found.

6.3.4.2 General Algorithmic Statement of GBD

Assuming that the problem (6.2) has a finite optimal value, Geoffrion (1972) stated the following general algorithm for **GBD**:

Step 1: Let an initial point $\boldsymbol{y}^1 \in \boldsymbol{Y} \cap \boldsymbol{V}$ (i.e., by fixing $\boldsymbol{y} = \boldsymbol{y}^1$, we have a feasible primal). Solve the resulting primal problem $P(\boldsymbol{y}^1)$ and obtain an optimal primal solution $\boldsymbol{x}^1$ and optimal multipliers; vectors λ^1, μ^1. Assume that you can find, somehow, the support function $\xi(y; \lambda^1, \mu^1)$ for the obtained multipliers λ^1, μ^1. Set the counters $k = 1$ for feasible and $l = 1$ for infeasible and the current upper bound $UBD = v(y^1)$. Select the convergence tolerance $\epsilon \geq 0$.

Step 2: Solve the relaxed master problem, **(RM)**:

$$\begin{aligned} \min_{y \in Y, \mu_B} \quad & \mu_B \\ s.t. \quad \mu_B &\geq \xi(y; \lambda^k, \mu^k), \quad k = 1, 2, \ldots, K \\ 0 &\geq \bar{\xi}(y; \bar{\lambda}^l, \bar{\mu}^l), \quad l = 1, 2, \ldots, \Lambda \end{aligned}$$

Let $(\hat{y}, \hat{\mu}_B)$ be an optimal solution of the above relaxed master problem. $\hat{\mu}_B$ is a lower bound on problem (6.2); that is, the current lower bound is $LBD = \hat{\mu}_B$. If $UBD - LBD \leq \epsilon$, then terminate.

Step 3: Solve the primal problem for $y = \hat{y}$, that is the problem $P(\hat{y})$. Then we distinguish two cases; feasible and infeasible primal:

Step 3a - Feasible Primal $P(\hat{y})$
The primal has $v(\hat{y})$ finite with an optimal solution $\hat{x}$ and optimal multiplier vectors $\hat{\lambda}, \hat{\mu}$. Update the upper bound $UBD = \min\{UBD, v(\hat{y})\}$. If $UBD - LBD \leq \epsilon$, then terminate. Otherwise, set $k = k + 1$, $\lambda^k = \hat{\lambda}$, and $\mu^k = \hat{\mu}$. Return to step 2, assuming we can somehow determine the support function $\xi(y; \lambda^{k+1}, \mu^{k+1})$.

Step3b - Infeasible Primal $P(\hat{y})$

The primal does not have a feasible solution for $y = \hat{y}$. Solve a feasibility problem (e.g., the l_1–minimization) to determine the multiplier vectors $\bar{\bar{\lambda}}, \bar{\bar{\mu}}$ of the feasibility problem.

Set $l = l + 1$, $\bar{\lambda}^l = \bar{\bar{\lambda}}$, and $\bar{\mu}^l = \bar{\bar{\mu}}$. Return to step 2, assuming we can somehow determine the support function $\bar{\xi}(y; \bar{\lambda}^{l+1}, \bar{\mu}^{l+1})$.

Remark 1 Note that a feasible initial primal is needed in step 1. However, this does not restrict the **GBD** since it is possible to start with an infeasible primal problem. In this case, after detecting that the primal is infeasible, step 3b is applied, in which a support function $\bar{\xi}$ is employed.

Remark 2 Note that step 1 could be altered, that is instead of solving the primal problem we could solve a continuous relaxation of problem (6.2) in which the y variables are treated as continuous bounded by zero and one:

$$\begin{aligned} \min_{x,y} \quad & f(x,y) \\ s.t. \quad & h(x,y) = 0 \\ & g(x,y) \leq 0 \\ & x \in X \\ & 0 \leq y \leq 1 \end{aligned} \tag{6.10}$$

If the solution of (6.10) is integral, then we terminate. If there exist fractional values of the y variables, then these can be rounded to the closest integer values, and subsequently these can be used as the starting y^1 vector with the possibility of the resulting primal problem being feasible or infeasible.

Remark 3 Note also that in step 1, step 3a, and step 3b a rather important assumption is made; that is, we can find the support functions ξ and $\bar{\xi}$ for the given values of the multiplier vectors (λ, μ) and $(\bar{\lambda}, \bar{\mu})$. The determination of these support functions cannot be achieved in general, since these are parametric functions of y and result from the solution of the inner optimization problems. Their determination in the general case requires a global optimization approach as the one proposed by (Floudas and Visweswaran, 1990; Floudas and Visweswaran, 1993). There exist however, a number of special cases for which the support functions can be obtained explicitly as functions of the y variables. We will discuss these special cases in the next section. If however, it is not possible to obtain explicitly expressions of the support functions in terms of the y variables, then assumptions need to be introduced for their calculation. These assumptions, as well as the resulting variants of **GBD** will be discussed in the next section. The point to note here is that the validity of lower bounds with these variants of **GBD** will be limited by the imposed assumptions.

Remark 4 Note that the relaxed master problem (see step 2) in the first iteration will have as a constraint one support function that corresponds to feasible primal and will be of the form:

$$\begin{aligned} \min_{y \in Y, \mu_B} \quad & \mu_B \\ s.t. \quad & \mu_B \geq \xi(y; \lambda^1, \mu^1) \end{aligned} \tag{6.11}$$

In the second iteration, if the primal is feasible and (λ^2, μ^2) are its optimal multiplier vectors, then the relaxed master problem will feature two constraints and will be of the form:

$$\begin{aligned} \min_{y \in Y, \mu_B} \quad & \mu_B \\ s.t. \quad & \mu_B \geq \xi(y; \lambda^1, \mu^1) \\ & \mu_B \geq \xi(y; \lambda^2, \mu^2) \end{aligned} \tag{6.12}$$

Note that in this case, the relaxed master problem (6.12) will have a solution that is greater or equal to the solution of (6.11). This is due to having the additional constraint. Therefore, we can see that the sequence of lower bounds that is created from the solution of the relaxed master problems is nondecreasing. A similar argument holds true in the case of having infeasible primal in the second iteration.

Remark 5 Note that since the upper bounds are produced by fixing the y variables to different 0-1 combinations, there is no reason for the upper bounds to satisfy any monotonicity property. If we consider however the updated upper bounds (i.e., $UBD = \min_k v(y^k)$), then the sequence for the updated upper bounds is monotonically nonincreasing since by their definition we always keep the best(least) upper bound.

Remark 6 The termination criterion for **GBD** is based on the difference between the updated upper bound and the current lower bound. If this difference is less than or equal to a prespecified tolerance $\varepsilon \geq 0$ then we terminate. Note though that if we introduce in the relaxed master integer cuts that exclude the previously found 0–1 combinations, then the termination criterion can be met by having found an infeasible master problem (i.e., there is no 0–1 combination that makes it feasible).

6.3.4.3 Finite Convergence of GBD

For formulation (6.2), Geoffrion (1972) proved finite convergence of the **GBD** algorithm stated in section 6.3.4.2, which is as follows:

Theorem 6.3.4 (Finite Convergence)
If C1, C2, C3 hold and **Y** *is a discrete set, then the* **GBD** *algorithm terminates in a finite number of iterations for any given* $\epsilon > 0$ *and even for* $\epsilon = 0$.

Note that in this case exact convergence can be obtained in a finite number of iterations.

6.3.5 Variants of GBD

In the previous section we discussed the general algorithmic statement of **GBD** and pointed out (see remark 3) a key assumption made with respect to the calculation of the support functions $\xi(y; \lambda, \mu)$ and $\bar{\xi}(y; \bar{\lambda}, \bar{\mu})$ from the feasible and infeasible primal problems, respectively. In this section, we will discuss a number of variants of **GBD** that result from addressing the calculation of the aforementioned support functions either rigorously for special cases or making assumptions that may not provide valid lower bounds in the general case.

6.3.5.1 Variant 1 of GBD, v1–GBD

This variant of **GBD** is based on the following assumption that was denoted by Geoffrion (1972) as Property **(P)**:

Theorem 6.3.5 (Property (P))
For every λ and $\mu \geq 0$, the infimum of $L(x, y, \lambda, \mu)$ with respect to $x \in X$ (i.e. the support $\xi(y; \lambda, \mu)$) can be taken independently of y so that the support function $\xi(y; \lambda, \mu)$ can be obtained explicitly with little or no more effort than is required to evaluate it at a single value of y. Similarly, the support function $\bar{\xi}(y; \bar{\lambda}, \bar{\mu})$, $(\bar{\lambda}, \bar{\mu}) \in \Lambda$ can be obtained explicitly.

Geoffrion (1972) identified the following two important classes of problems where Property (P) holds:

Class 1: f, h, g are linearly separable in x and y.

Class 2: Variable factor programming

Geromel and Belloni (1986) identified a similar class to variable factor programming that is applicable to the unit commitment of thermal systems problems.

In class 1 problems, we have

$$\begin{aligned} f(x,y) &= f_1(x) + f_2(y), \\ h(x,y) &= h_1(x) + h_2(y), \\ g(x,y) &= g_1(x) + g_2(y). \end{aligned}$$

In class 2 problems, we have

$$\begin{aligned} f(x,y) &= -\sum_i f_i(x^i)y_i, \\ g(x,y)_j &= \sum_i x^i y_i - c. \end{aligned}$$

In Geromel and Belloni (1986) problems, we have

$$\begin{aligned} f(x,y) &= \sum_k \sum_i f_i(x_i(k))y_i + \sum_i g_i(y_i), \\ g(x,y)_j &= -\sum_i x_i(k)y_i - L(k). \end{aligned}$$

In the sequel, we will discuss the **v1–GBD** for class 1 problems since this by itself defines an interesting mathematical structure for which other algorithms (e.g., Outer Approximation) has been developed.

v1–GBD under separability

Under the separability assumption, the support functions $\xi(y; \lambda^k, \mu^k)$ and $\bar{\xi}(y; \bar{\lambda}^l, \bar{\mu}^l)$ can be obtained as explicit functions of y since

$$
\begin{aligned}
\xi(y;\lambda^k,\mu^k) &= \min_{x\in X} L(x,y,\lambda^k\mu^k) \\
&= \min_{x\in X} \{f(x,y)) + \lambda^{k^T} h(x,y) + \mu^{kT} g(x,y)\} \\
&= \min_{x\in X} \{f_1(x) + f_2(y) + \lambda^{kT}(h_1(x) + h_2(y)) + \mu^{kT}(g_1(x) + g_2(y))\} \\
&= f_2(y) + \lambda^{kT} h_2(y) + \mu^{kT} g_2(x) + \min_{x\in X}[f_1(x) + \lambda^{kT} h_1(x) + \mu^{kT} g_1(x)].
\end{aligned}
$$

Remark 1 Note that due to separability we end up with an explicit function of y and a problem only in x that can be solved independently.

Similarly, the support function $\bar{\xi}(y;\bar{\lambda}^l,\bar{\mu}^l)$ is

$$
\begin{aligned}
\bar{\xi}(y;\bar{\lambda}^l,\bar{\mu}^l) &= \min_{x\in X} \bar{L}(x,y,\bar{\lambda}^l,\bar{\mu}^l) \\
&= \min_{x\in X} \left\{\bar{\lambda}^{l^T} h(x,y) + \bar{\mu}^{l^T} g(x,y)\right\} \\
&= \min_{x\in X} \left\{\bar{\lambda}^{l^T} (h_1(x,y) + h_2(x,y)) + \bar{\mu}^{l^T} (g_1(x,y) + g_2(x,y))\right\} \\
&= \bar{\lambda}^{l^T} h_2(y) + \bar{\mu}^{l^T} g_2(y) + \min_{x\in X} \left[\bar{\lambda}^{l^T} h_1(x) + \bar{\mu}^{l^T} g_1(x)\right].
\end{aligned}
$$

Remark 2 Note that to solve the independent problems in x, we need to know the multiplier vectors (λ^k,μ^k) and $(\bar{\lambda}^l,\bar{\mu}^l)$ from feasible and infeasible primal problems, respectively.

Under the separability assumption, the primal problem for fixed $y = y^k$ takes the form

$$
\begin{aligned}
\min_{x\in X} \quad & f_1(x) + f_2(y^k) \\
s.t. \quad & h_1(x) = -h_2(y^k) \\
& g_1(x) \le -g_2(y^k)
\end{aligned}
$$

Now, we can state the algorithmic procedure for the **v1–GBD** under the separability assumption.

v1–GBD Algorithm

Step 1: Let an initial point $y^1 \in Y \cap V$. Solve the primal $P(y^1)$ and obtain an optimal solution x^1, and multiplier vectors λ^1,μ^1. Set the counters $k = 1, l = 1$, and $UBD = v(y^1)$. Select the convergence tolerance $\epsilon \ge 0$.

Step 2: Solve the relaxed master problem:

$$
\begin{aligned}
\min_{y\in Y,\mu_B} \quad & \mu_B \\
s.t. \quad & \mu_B \ge f_2(y) + \lambda^{k^T} h_2(y) + \mu^{k^T} g_2(y) + L_1^k, \quad k = 1,2,\ldots,K \\
& 0 \ge \mu_B \bar{\lambda}^{l^T} h_2(y) + \bar{\mu}^{l^T} g_2(y) + L_1^l, \quad l = 1,2,\ldots,\Lambda
\end{aligned}
$$

$$\text{where} \quad L_1^k = \min_{x \in X} \left[f_1(x) + \lambda^{k^T} h_1(x) + \mu^{k^T} g_1(x) \right]$$
$$\bar{L}_1^k = \min_{x \in X} \left[f_1(x) + \bar{\lambda}^{l^T} h_1(x) + \bar{\mu}^{l^T} g_1(x) \right]$$

are solutions of the above stated independent problems.

Let $(\hat{y}, \hat{\mu}_B)$ be an optimal solution. $\hat{\mu}_B$ is a lower bound, that is $LBD = \hat{\mu}_B$. If $UBD - LBD \leq \epsilon$, then terminate.

Step 3: As in section 6.3.4.2.

Remark 3 Note that if in addition to the separability of x and y, we assume that y participates linearly (i.e., conditions for Outer Approximation algorithm), then we have

$$\begin{aligned} f_2(y) &= c^T y, \\ h_2(y) &= Ay, \\ g_2(y) &= By, \end{aligned}$$

in which case, the relaxed master problem of step 2 of **v1–GBD** will be a linear 0–1 programming problem with an additional scalar μ_B, which can be solved with available solvers (e.g., CPLEX, ZOOM, SCICONIC).

If the y variables participate separably but in a nonlinear way, then the relaxed master problem is of 0–1 nonlinear programming type.

Remark 4 Note that due to the strong duality theorem we do not need to solve the problems for $L_1^k, \bar{L}_1^l$ since their optimum solutions are identical to the ones of the corresponding feasible and infeasible primal problems with respect to x, respectively.

Illustration 6.3.1 This example is a modified version of example 1 of Kocis and Grossmann (1987) and can be stated as follows:

$$\begin{aligned} \min_{x_1, y} \quad & -y + 2x_1 - \ln(0.5x_1) \\ \text{subject to} \quad & -x_1 - \ln(0.5x_1) + y \leq 0 \\ & 0.5 \leq x_1 \leq 1.4 \\ & y = \{0, 1\} \end{aligned}$$

Note that it features separability in x and y and linearity in y.

$$\begin{aligned} f_1(x) &= 2x_1 - \ln(0.5x_1), \\ f_2(y) &= -y, \\ g_1(x) &= -x_1 - \ln(0.5x_1), \\ g_2(y) &= y. \end{aligned}$$

Also note that $f_1(x), g_1(x)$ are convex functions in x_1, and hence the required convexity conditions are satisfied.

Based on the previously presented analysis for the **v1–GBD** under the separability assumption, we can now formulate the relaxed master problem in an explicit form.

Relaxed Master Problem

$$\begin{aligned} \min_{y,\mu_B} \quad & \mu_B \\ \text{subject to} \quad & \mu_B \geq -y + \mu^k y + L_1^k, \quad k = 1, 2, \ldots, K \\ & 0 \geq \bar{\mu}^l y + \bar{L}_1^l, \quad l = 1, 2, \ldots, L \end{aligned}$$

$$\text{where} \quad \begin{aligned} L_1^k &= \min_{0.5 \leq x_1 \leq 1.4} 2x_1 - \ln(0.5x_1) + \mu^k\left(-x1 - \ln(0.5x_1)\right) \\ L_1^l &= \min_{0.5 \leq x_1 \leq 1.4} \mu^l\left(-x1 - \ln(0.5x_1)\right) \end{aligned}$$

Now we can apply the **v1–GBD** algorithm.

Step 1: Select $y^1 = 0$.

Solve the following primal problem $P(y^1)$.

$$\begin{aligned} \min_{x_1} \quad & 2x_1 - 2\ln(0.5x_1) \\ \text{subject to} \quad & -x_1 - \ln(0.5x_1) \leq 0 \\ & 0.5 \leq x_1 \leq 1.4 \end{aligned}$$

which has as solution

$$\begin{aligned} x_1 &= 0.353, \\ \mu^1 &= 0.381, \end{aligned}$$

and the upper bound is $UBD = 2.558$.

Step 2:

$$\begin{aligned} L_1^1 &= \min_{0.5 \leq x_1 \leq 1.4} 2x_1 - \ln(0.5x_1) + 0.381\left(-x_1 - \ln(0.5x_1)\right) \\ &= 2.558. \end{aligned}$$

Note that we do not not need to solve for L_1^k since due to strong duality its solution is identical to the one of the corresponding primal problem $P(y^1)$.

Then, the relaxed master problem is of the form:

$$\begin{aligned} \min_{y,\mu_B} \quad & \mu_B \\ \text{subject to} \quad & \mu_B \geq -y + 0.381y + 2.558 \\ & y = 0, 1 \end{aligned}$$

Its solution is $y^2 = 1$ and the lower bound is

$$LBD = 1.939.$$

Step 3: Solve the Primal for $y^2 = 1,\ P(y^2)$ which has as solution:

$$\begin{aligned} x_1 &= 1.375, \\ \text{objective} &= 2.124, \\ \mu &= 0.73684. \end{aligned}$$

The new upper bound is

$$UBD = \min(2.558, 2.124) = 2.124,$$

and this is the optimal solution since we have examined all 0–1 combinations.

6.3.5.2 Variant 2 of GBD, v2–GBD

This variant of **GBD** is based on the assumption that we can use the optimal solution x^k of the primal problem $P(y^k)$ along with the multiplier vectors for the determination of the support function $\xi(y; \lambda^k, \mu^k)$.

Similarly, we assume that we can use the optimal solution of the feasibility problem (if the primal is infeasible) for the determination of the support function $\bar{\xi}(y; \lambda^k, \mu^k)$.

The aforementioned assumption fixes the x vector to the optimal value obtained from its corresponding primal problem and therefore eliminates the inner optimization problems that define the support functions. It should be noted that fixing x to the solution of the corresponding primal problem may not necessarily produce valid support functions in the sense that there would be no theoretical guarantee for obtaining lower bounds to solution of (6.2) can be claimed in general.

v2–GBD Algorithm

The **v2–GBD** algorithm can be stated as follows:

Step 1: Let an initial point $y^1 \in Y \cap V$.
Solve the primal problem $P(y^1)$ and obtain an optimal solution x^1 and multiplier vectors λ^1, μ^1. Set the counters $k = 1, l = 1$, and $UBD = v(y^1)$ Select the convergence tolerance $\epsilon \geq 0$.

Step 2: Solve the relaxed master problem:

$$\begin{aligned} \min_{y \in Y, \mu_B} \quad & \mu_B \\ s.t. \quad \mu_B &\geq L(x^k, y, \lambda^k, \mu^k), \quad k = 1, 2, \ldots, K \\ 0 &\geq \bar{L}(x^l, y, \bar{\lambda}^l, \bar{\mu}^l), \quad l = 1, 2, \ldots, \Lambda \end{aligned}$$

$$\text{where} \quad L(x^k, y, \lambda^k, \mu^k) = f(x^k, y) + {\lambda^k}^T h(x^k, y) + {\mu^k}^T g(x^k, y)$$
$$\bar{L}(\bar{x}^l, y, \bar{\lambda}^l, \bar{\mu}^k) = {\bar{\lambda}^k}^T h(x^l, y) + {\bar{\mu}^k}^T g(x^l, y)$$

are the Lagrange functions evaluated at the optimal solution x^k of the primal problem.

Let $(\hat{y}, \hat{\mu}_B)$ be an optimal solution. $\hat{\mu}_B$ is a lower bound, that is, $LBD = \hat{\mu_B}$. If $UBD - LBD \leq \epsilon$, then terminate.

Step 3: As in section 6.3.4.2.

Remark 1 Note that since $y \in Y = \{0 - 1\}$, the master problem is a 0–1 programming problem with one scalar variable μ_B. If the y variables participate linearly, then it is a 0–1 linear problem which can be solved with standard branch and bound algorithms. In such a case, we can introduce integer cuts of the form:

$$\sum_{i \in B} y_i - \sum_{i \in NB} y_i \leq |B| - 1,$$
$$\text{where} \quad B = \{i : y_i = 1\},$$
$$NB = \{i : y_i = 0\},$$
$$|B| \quad \text{is the cardinality of B},$$

which eliminate the already found 0–1 combinations. If we employ such a scheme, then an alternative termination criterion is that of having infeasible relaxed master problems. This of course implies that all 0–1 combinations have been considered.

Remark 2 It is of considerable interest to identify the conditions which if satisfied make the assumption in **v2–GBD** a valid one. The assumption in a somewhat different restated form is that:

$$\xi(y; \lambda^k, \mu^k) = \min_{x \in X} L(x, y, \lambda^k, \mu^k) \geq L(x^k, y, \lambda^k, \mu^k), \quad k = 1, 2, \ldots, K,$$
$$\bar{\xi}(y; \bar{\lambda}^l, \mu^l) = \min_{x \in X} \bar{L}(x, y, \bar{\lambda}^l, \bar{\mu}^l) \geq \bar{L}(x^l, y, \bar{\lambda}^l, \bar{\mu}^l), \quad l = 1, 2, \ldots, \Lambda,$$

that is, we assume that the Lagrange function evaluated at the solution of the corresponding primal are valid underestimators of the inner optimization problems with respect to $x \in X$.

Due to condition C1 the Lagrange functions $L(x, y, \lambda^k, \mu^k)$, $\bar{L}(x, y, \bar{\lambda}^l, \bar{\mu}^l)$ are convex in x for each fixed y since they are linear combinations of convex functions in x.

$L(x, y, \lambda^k, \mu^k)$, $\bar{L}(\bar{x}^l, y, \bar{\lambda}^l, \bar{\mu}^l)$ represent local linearizations around the points x^k and $\bar{x}^k$ of the support functions $\xi(y; \lambda^k, \mu^k)$, $\bar{\xi}(y; \bar{\lambda}^l, \mu^l)$, respectively. Therefore, the aforementioned assumption is valid if the projected problem $v(y)$ is convex in y. If, however, the projected problem $v(y)$ is nonconvex, then the assumption does not hold, and the algorithm may terminate at a local (nonglobal) solution or even at a nonstationary point. This analysis was first presented by Floudas and Visweswaran (1990) and later by Sahinidis and Grossmann (1991), and Bagajewicz and Manousiouthakis (1991). Figure 6.1 shows the case in which the assumption is valid, while Figure 6.2 shows a case for which a local solution or a nonstationary point may result.

Note that in the above analysis we did not assume that $Y = \{0, 1\}^q$, and hence the argument is applicable even when the y–variables are continuous. In fact, Figures 6.1 and 6.2 represent continuous y variables.

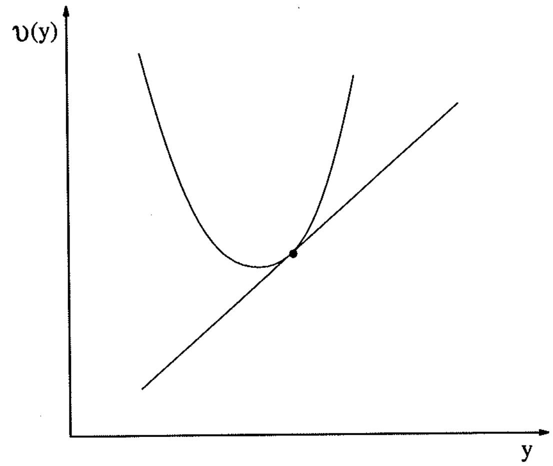

Figure 6.1: Valid support

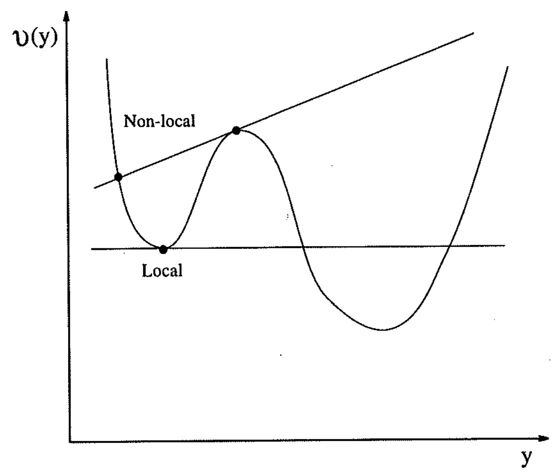

Figure 6.2: Invalid support

Remark 3 It is also very interesting to examine the validity of the assumption made in **v2–GBD** under the conditions of separability of x and y and linearity in y (i.e., **OA** conditions). In this case we have:

$$\begin{aligned} f(x,y) &= \mathbf{c}^T y + f_1(x), \\ h(x,y) &= Ay + h_1(x), \\ g(x,y) &= By + g_1(x). \end{aligned}$$

Then the support function for feasible primal becomes

$$\begin{aligned} \xi(y;\lambda^k,\mu^k) &= \mathbf{c}^T y + \lambda^{k^T}(Ay) + \mu^{k^T}(By) \\ &\quad + \min_{x \in X} f_1(x) + \lambda^{k^T} h_1(x) + \mu^{k^T} g_1(x), \end{aligned}$$

which is linear in y and hence convex in y. Note also that since we fix $x = x^k$, the $\min_{x \in X}$ is in fact an evaluation at x^k. Similarly the case for $\bar{\xi}(y;\bar{\lambda}^k,\mu^k)$ can be analyzed.

Therefore, the assumption in **v2–GBD** holds true if separability and linearity hold which covers also the case of linear 0–1 y variables. This way under conditions C1, C2, C3 the **v2–GBD** determined the global solution for separability in x and y and linearity in y problems.

Illustration 6.3.2 This example is taken from Sahinidis and Grossmann (1991) and has three 0–1 variables.

$$\begin{aligned} \min \quad y_1 + y_2 &+ y_3 + 5x^2 \\ s.t. \quad 3x - y_1 - y_2 &\leq 0 \\ -x + 0.1y_2 + 0.25y_3 &\leq 0 \\ y_1 + y_2 + y_3 &\geq 2 \\ y_1 + y_2 + 2(y_3 - 1) &\geq 0 \\ 0.2 \leq x &\leq 1 \\ y_1, y_2, y_3 &= 0, 1 \end{aligned}$$

Note that the third and fourth constraint have only 0–1 variables and hence can be moved directly to the relaxed master problem.

Also, note that this example has separability in x and y, linearity in y, and convexity in x for fixed y. Thus, we have

$$\begin{aligned} \xi(y;\lambda^k,\mu^k) &= y_1 + y_2 + y_3 + \mu_1^k(-y_1 - y_2) + \mu_2^k(0.1y_2 + 0.25y_3), \\ &\quad + \left[5x^{k^2} + \mu_1^k(3x^k) + \mu_2^k(-x^k)\right] \\ \bar{\xi}(y;\bar{\lambda}^k,\bar{\mu}^k) &= +\mu_1^l(-y_1 - y_2) + \mu_2^l(0.1y_2 + 0.25y_3) + [\bar{\mu}_1^l(3x^l) + \bar{\mu}_2^l(-x^l)]. \end{aligned}$$

Iteration 1

Step 1: Set $(y_1, y_2, y_3) = (1, 1, 1)$
The primal problem becomes

$$\begin{aligned} \min \quad 3 &+ 5x^2 \\ s.t. \quad 3x - 2 &\leq 0 \\ -x + 0.35 &\leq 0 \\ 0.2 \leq x &\leq 1 \end{aligned}$$

and its solution is

$$\begin{aligned} x^1 &= 0.35, \\ \mu_1^1 &= 0, \\ \mu_2^1 &= 3.5, \end{aligned}$$

with objective equal to the $UBD = 3.6125$.

Step 2: The relaxed master problem is

$$\begin{aligned} \min_{y_1, y_2, y_3, \mu_B} \quad & \mu_B \\ s.t. \quad \mu_B &\geq y_1 + y_2 + y_3 + 0\,(-y_1 - y_2)\,3.5\,(0.1y_2 + 0.25y_3) + 0.6125 \\ y_1 + y_2 + y_3 &\geq 2 \\ y_1 + y_2 + 2(y_3 - 1) &\geq 0 \\ y_1, y_2, y_3 &= 0, 1 \end{aligned}$$

which has as solution $y^2 = (1, 1, 0)$ and $\mu_B = LBD = 1.7375$. Since $UBD - LBD = 3.6125 - 1.735 = 1.8775$, we continue with $y = y^2$. Note that $5(0.35)^2 + (0)(30.35) + (3.5)(-0.35) = -0.6125$.

Iteration 2

Step 1: Set $y^2 = (1, 1, 0)$
Solve the primal $P(y^2)$

$$\begin{aligned} \min \quad 2 &+ 5x^2 \\ s.t. \quad 3x - 2 &\leq 0 \\ -x + 0.1 &\leq 0 \\ 0.2 \leq x &\leq 1 \end{aligned}$$

and its solution is

$$\begin{aligned} x^2 &= 0.2, \\ \mu_1^2 &= 0, \\ \mu_2^2 &= 0, \end{aligned}$$

and the objective function is 2.2. The updated upper bound is $UBD = \min(3.6125, 2.2) = 2.2$.

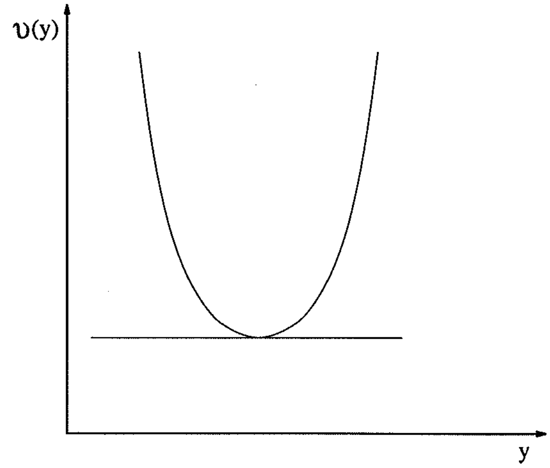

Figure 6.3: Termination of GBD in one iteration

Step 2: The relaxed master problem is:

$$\begin{aligned} & \min \quad \mu_B \\ s.t. \quad \mu_B &\geq y_1 + y_2 + y_3 + 3.5\,(0.1y_2 + 0.25y_3) - 0.6125 \\ \mu_B &\geq y_1 + y_2 + y_3 + 0.2 \\ y_1 + y_2 + y_3 &\geq 2 \\ y_1 + y_2 + 2(y_3 - 1) &\geq 0 \\ y_1, y_2, y_3 &= 0, 1 \end{aligned}$$

which has as solution $(y_1, y_2, y_3) = (1, 1, 0)$ and $\mu_B = LBD = 2.2$. Since $UBD - LBD = 0$, we terminate with $(y_1, y_2, y_3) = (1, 1, 0)$ as the optimal solution.

Remark 4 Note that if we had selected as the starting point the optimal solution; that is $(1, 1, 0)$, then the **v2–GBD** would have terminated in one iteration. This can be explained in terms of Remark 3. Since $v(y)$ is convex, then the optimal point corresponds to the global minimum and the tangent plane to this minimum provides the tightest lower bound which by strong duality equals the upper bound. This is illustrated in Figure 6.3.

6.3.5.3 Variant 3 of GBD, v3–GBD

This variant was proposed by Floudas *et al.* (1989) and denoted as Global Optimum Search **GOS** and was applied to continuous as well as 0–1 set Y. It uses the same assumption as the one in **v2–GBD** but *in addition* assumes that

(i) $f(x,y), g(x,y)$ are convex functions in y for every fixed x, and

(ii) $h(x,y)$ are linear functions in y for every x.

This additional assumption was made so as to create special structure not only in the primal but also in the relaxed master problem. The type of special structure in the relaxed master problem has to do with its convexity characteristics.

The basic idea in **GOS** is to select the x and y variables in a such a way that the primal and the relaxed master problem of the **v2–GBD** satisfy the appropriate convexity requirements and hence attain their respective global solutions.

We will discuss **v3–GBD** first under the separability of x and y and then for the general case.

v3–GBD with separability

Under the separability assumption we have

$$\begin{aligned} f(x,y) &= f_1(x) + f_2(y), \\ h(x,y) &= h_1(x) + h_2(y), \\ g(x,y) &= g_1(x) + g_2(y). \end{aligned}$$

The additional assumption that makes **v3–GBD** different than **v2–GBD** implies that

(i) $f_2(y), g_2(y)$ are convex in y, and

(ii) $h_2(y)$ are linear in y.

Then, the relaxed master problem will be

$$\begin{aligned} \min_{y,\mu_B} \quad & \mu_B \\ s.t. \quad \mu_B \geq\ & f_2(y) + \lambda^{k^T} h_2(y) + \mu^{k^T} g_2(y) \\ & + \left[f_1(x^k) + \lambda^{k^T} h_1(x^k) + \mu^{k^T} g_1(x^k) \right], \quad k = 1,2,\ldots,K \\ 0 \geq\ & \bar{\lambda}^{l^T} h_2(y) + \bar{\mu}^{l^T} g_2(y) + \left[\bar{\lambda}^{l^T} h_1(\bar{x}^l) + \bar{\mu}^{l^T} g_1(\bar{x}^l) \right], \quad l = 1,2,\ldots,L \end{aligned}$$

Remark 1 Note that the additional assumption makes the problem convex in y if y represent continuous variables. If $y \in Y = \{0,1\}^q$, and the y–variables participate linearly (i.e. f_2, g_2 are linear in y), then the relaxed master is convex. Therefore, this case represents an improvement over **v3–GBD**, and application of **v3–GBD** will result in valid support functions, which implies that the global optimum of (6.2) will be obtained.

v3–GBD without separability

The Global Optimum Search **GOS** aimed at exploiting and invoking special structure for nonconvex nonseparable problems of the type (6.2).

$$\min \quad f(x,y)$$

$$\begin{aligned} \text{s.t.} \quad h(x,y) &= 0 \\ g(x,y) &\leq 0 \\ x &\in X \subseteq \Re^n \\ y &\in Y \subseteq \Re^q \end{aligned}$$

under the conditions C1, C2, C3 and the additional condition:

(i) $f(x,y)$, $g(x,y)$ are convex functions in y for every fixed x,

(ii) $h(x,y)$ are linear functions in y for every x;

so that both the primal and the relaxed problems attain their respective global solutions.

Remark 2 Note that since x and y are not separable, then the **GOS** cannot provide theoretically valid functions in the general case, but only if the $v(y)$ is convex (see **v2–GBD**).

Despite this theoretical limitation, it is instructive to see how, for $Y \subseteq \Re^n$, the convex primal and relaxed master problems are derived. This will be illustrated in the following.

Illustration 6.3.3 This is an example taken from Floudas *et al.* (1989).

$$\begin{aligned} \min \quad & -12x_1 - 7x_2 + x_2^2 \\ \text{s.t.} \quad & -2x_1^4 - x_2 + 2 = 0 \\ & 0 \leq x_1 \leq 2 \\ & 0 \leq x_2 \leq 3 \end{aligned}$$

Note that the objective function is convex since it has linear and positive quadratic terms. The only nonlinearities come from the equality constraint. By introducing three new variables w_1, w_2, w_3, and three equalities:

$$\begin{aligned} w_1 - x_1 &= 0, \\ w_2 - x_1 w_1 &= 0, \\ w_3 - x_1 w_2 &= 0, \end{aligned}$$

we can write an equivalent formulation of

$$\begin{aligned} \min \quad & -12x_1 - 7x_2 + x_2^2 \\ \text{s.t.} \quad & -2w_3 x_1 - x_2 + 2 = 0 \\ & w_1 - x_1 = 0 \\ & w_2 - x_1 w_1 = 0 \\ & w_3 - x_1 w_2 = 0 \\ & 0 \leq x_1 \leq 2 \\ & 0 \leq x_2 \leq 3 \end{aligned}$$

Note that if we select as

$$\begin{aligned} y &= x_1, \\ x &= (x_2, w_1, w_2, w_3), \end{aligned}$$

all the imposed convexity conditions are satisfied and hence the primal and the relaxed master problems are convex and attain their respective global solutions.

The primal problem for $y = y^k$ is

$$\begin{aligned} \min \quad & -12y^k - 7x_2 + x_2^2 \\ \text{s.t.} \quad & -2w_3y^k - x_2 + 2 = 0 \\ & w_1 - y^k = 0 \\ & w_2 - y^k w_1 = 0 \\ & w_3 - y^k w_2 = 0 \\ & 0 \leq x_2 \leq 3 \end{aligned}$$

The relaxed master problem is

$$\begin{aligned} \min_{y,\mu_B} \quad & \mu_B \\ \text{s.t.} \quad & \mu_B \geq L(x^k, y, \lambda^k, \mu^k), \quad k = 1,2,\ldots,K \\ & 0 \geq \bar{L}(\bar{x}^l, y, \bar{\lambda}^l, \bar{\mu}^l), \quad l = 1,2,\ldots,L \\ & 0 \leq y \leq 2 \end{aligned}$$

$$\begin{aligned} \text{where} \quad L(x^k, y, \lambda^k, \mu^k) = \; & -12y - 7x_2^k + (x_2^k)^2 \\ & + \lambda_1^k(-2w_3^k y - x_2^k + 2) \\ & + \lambda_2^k(w_1^k - y) \\ & + \lambda_3^k(w_2^k - y w_1^k) \\ & + \lambda_4^k(w_3^k - y w_2^k); \end{aligned}$$

$$\begin{aligned} \bar{L}(\bar{x}^l, y, \bar{\lambda}^l, \bar{\mu}^l) = \; & \bar{\lambda}_1^l(-2\bar{w}_3^l y - \bar{x}_2^l + 2) \\ + \; & \bar{\lambda}_2^l(\bar{w}_1^l - y) \\ + \; & \bar{\lambda}_3^l(\bar{w}_2^l - y\bar{w}_1^l) \\ + \; & \bar{\lambda}_4^l(\bar{w}_3^l - y\bar{w}_2^l). \end{aligned}$$

Remark 3 The primal is convex in x, while the relaxed master is linear in y.

Application of **v3–GBD** from several starting points determines the global solution which is:

$$\begin{aligned} y &= 0.718, \\ x_2 &= 1.47, \\ \text{Objective} &= -16.7389. \end{aligned}$$