Journal of Global Optimization **20:** 1–31, 2001.

Global Optimization of Nonlinear Bilevel Programming Problems

ZEYNEP H. GÜMÜŞ and CHRISTODOULOS A. FLOUDAS*
Department of Chemical Engineering, Princeton University, Princeton, N.J. 08544-5963, USA (e-mail: floudas@titan.princeton.edu)

(Accepted 17 November 2000)

Abstract. A novel technique that addresses the solution of the general nonlinear bilevel programming problem to global optimality is presented. Global optimality is guaranteed for problems that involve twice differentiable nonlinear functions as long as the linear independence constraint qualification condition holds for the inner problem constraints. The approach is based on the relaxation of the feasible region by convex underestimation, embedded in a branch and bound framework utilizing the basic principles of the deterministic global optimization algorithm, $\boldsymbol{\alpha}$**BB** [2, 4, 5, 11]. Epsilon global optimality in a finite number of iterations is theoretically guaranteed. Computational studies on several literature problems are reported.

Key words: Bilevel programming, Bilevel optimization, Twice-continuously differentiable, Global optimization, Bilevel nonlinear, Nonconvex, Mixed integer nonlinear optimization

1. Introduction

The bilevel programming problem, BLPP, is an optimization problem that is constrained by another optimization problem. This mathematical programming model arises when two independent decision makers, ordered within a hierarchical structure, have conflicting objectives. The decision maker at the lower level has to optimize her objective under the given parameters from the upper level decision maker, who, in return, with complete information on the possible reactions of the lower, selects the parameters so as to optimize her own objective. In this sense, the BLPP can be perceived as a static Stackelberg game [58, 65] with two independent decision makers. The decision maker with the upper level objective, $F(\mathbf{x}, \mathbf{y})$ takes the lead, and chooses her decision vector $\mathbf{x}$. The decision maker with lower level objective, $f(\mathbf{x}, \mathbf{y})$, reacts accordingly by choosing her decision vector $\mathbf{y}$ to optimize her objective, parameterized in $\mathbf{x}$. Note that the upper level decision maker is limited to influencing, rather than controlling, the lower level's outcome.

* Author to whom all correspondence should be addressed

1.1. PROBLEM DEFINITION

The general formulation of the BLPP is as follows:

$$
\begin{aligned}
\min_{\mathbf{x}} \quad & F(\mathbf{x}, \mathbf{y}) \\
\text{s.t.} \quad & \\
& \mathbf{G}(\mathbf{x}, \mathbf{y}) \leqslant 0 \\
& \mathbf{H}(\mathbf{x}, \mathbf{y}) = 0 \\
& \min_{\mathbf{y}} f(\mathbf{x}, \mathbf{y}) \\
& \text{s.t.} \\
& \qquad \mathbf{g}(\mathbf{x}, \mathbf{y}) \leqslant 0 \\
& \qquad \mathbf{h}(\mathbf{x}, \mathbf{y}) = 0 \\
& \mathbf{x} \in X \subset R^{n_1}, \quad \mathbf{y} \in Y \subset R^{n_2}
\end{aligned}
\tag{1}
$$

where $f, F : R^{n_1} \mathrm{x} R^{n_2} \rightarrow R$, $\mathbf{g} = [g_1, \dots, g_J] : R^{n_1} \mathrm{x} R^{n_2} \rightarrow R^J$, $\mathbf{G} = [G_1, \dots, G_{J'}] : R^{n_1} \mathrm{x} R^{n_2} \rightarrow R^{J'}$, $\mathbf{h} = [h_1, \dots, h_I] : R^{n_1} \mathrm{x} R^{n_2} \rightarrow R^I$, $\mathbf{H} = [H_1, \dots, H_{I'}] : R^{n_1} \mathrm{x} R^{n_2} \rightarrow R^{I'}$.

1.2. BACKGROUND

The BLPP model first appeared in a paper by Bracken and McGill [20], on the allocation of resources and weapons to optimize offense and defense simultaneously. However, Candler and Norton [23] have been the first to use the terms ‘bilevel’ or ‘multilevel’ while describing a development policy problem. Since then, the BLPP model has been employed in many and diverse areas that require hierarchical decision making. For example, centralized economic planning involves resource distribution through government levels [25], agricultural credit distribution [52], electric utility pricing and planning [38, 45] and tax-credit determination [16] problems that are naturally formulated as BLPPs. In civil engineering, extensive research has been conducted to solve the bilevel transportation network design problem [18, 44, 67]. In chemical engineering, BLPP applications involve chemical process design with equilibrium [26, 27, 36], plant design under uncertainty [39], flexibility analysis [35] and process design with controllability issues [21] problems. For comprehensive literature reviews and applications, the reader is directed to [43, 50, 51, 61].

Given that the BLPP applications are many and diverse, effective solution algorithms are of critical importance. The linear BLPP has the favorable property that the solution occurs at an extreme point of the feasible set, that can be exploited by enumeration techniques [19, 24]. However, this condition does not hold for the nonlinear BLPP.

The conventional solution approach to the nonlinear BLPP is to transform the original two level problem into a single level one by replacing the lower level optimization problem with the set of equations that define its Karush–Kuhn–Tucker, KKT, optimality conditions. However, the KKT optimality conditions are necessary and sufficient for defining the optimum of the inner level problem only under convexity conditions and a first order constraint qualification. When the inner problem constraints are nonconvex, the KKT conditions are only *necessary*. Consequently, local or even suboptimal solutions may be obtained.

A further difficulty arises in locating the global optimum of the resulting single level problem after the KKT transformation. The bilinear nature of complementarity conditions introduce nonconvexities even if the original problem is linear. Furthermore, when the inner problem is nonlinear, the equations that define the stationarity constraints are also nonconvex. Hence, even if the KKT conditions are *necessary and sufficient* for the inner problem, the global optimality of the transformed single level problem can not be guaranteed unless a global optimization algorithm is introduced. These difficulties related with the KKT-type solution approaches, which are the most efficient and widely used methods for the solution of the BLPP, confine them to only local solutions when nonlinearities are involved. On the other hand, many hierarchical systems typically involve nonlinear equations, and obtaining their global minimum may be of critical importance in decision making. Such is the case in process design under thermodynamical equilibrium and design under uncertainty problems. Therefore, it is extremely desirable to develop a technique that can locate the global optimum of nonlinear bilevel programming problems. However, to date, there have been relatively few studies on the global optimization of general nonlinear bilevel optimization problems. Furthermore, it is worth noting that there do not exist any rigorous approaches for bilevel nonlinear optimization problems in the open literature.

We have developed a novel global optimization technique for solving bilevel optimization problems that uses KKT optimality conditions, but can overcome their limitations by a systematic and rigorous procedure. Global optimality is guaranteed for general nonlinear bilevel optimization problems that may involve twice differentiable nonconvex nonlinear functions.

First, the KKT transformation will be described and the basic difficulties associated with obtaining a global solution using the current methods will be discussed. Next, the key properties of our proposed method for the solution of the general bilevel nonlinear optimization problems will be presented and the corresponding algorithmic procedure will be outlined. Finally, computational studies on several examples from the literature will be presented.

2. Theory

2.1. SET DEFINITIONS AND PROPERTIES

There are several set definitions related to BLPP solution approaches. The set

$$\Omega = \{(\mathbf{x}, \mathbf{y}) : \mathbf{G}(\mathbf{x}, \mathbf{y}) \leqslant 0, \mathbf{H}(\mathbf{x}, \mathbf{y}) = 0, \mathbf{g}(\mathbf{x}, \mathbf{y}) \leqslant 0, \mathbf{h}(\mathbf{x}, \mathbf{y}) = 0\}$$

defines the relaxed BLPP feasible set. If there are no $(\mathbf{x}, \mathbf{y}) \in \Omega$, then the BLPP is infeasible.

The feasible set of the inner problem is parametric in terms of the decision variables $\mathbf{x}$ of the outer problem, defined for every $\mathbf{x} \in X$ as:

$$\Omega(\mathbf{x}) = \{\mathbf{y} : \mathbf{y} \in Y, \mathbf{h}(\mathbf{x}, \mathbf{y}) = 0, \mathbf{g}(\mathbf{x}, \mathbf{y}) \leqslant 0\}$$

Thus, for every $\mathbf{x} \in X$, the set:

$$RR(\mathbf{x}) = \{\mathbf{y} \in \operatorname{argmin} f(\mathbf{x}, \mathbf{y}) : \mathbf{y} \in \Omega(\mathbf{x})\}$$

defines the inner problems rational reaction set. Hence, the BLPP feasible set (the inducable region) is defined as:

$$IR = \{(\mathbf{x}, \mathbf{y}) : (\mathbf{x}, \mathbf{y}) \in \Omega, \mathbf{y} \in RR(\mathbf{x})\}.$$

The solution of even the linear BLPP is an NP-hard problem [13, 17, 41] and furthermore, the BLPP is strongly NP-hard [37]. For studies on complexity issues regarding BLPPs, see [51, 53].

Note that at certain outer parameter values, the inner problem may have multiple optima, while the outer problem will be optimum only at specific inner variable values. When this situation arises, the optimum of the BLPP is achieved only if the inner optimizer cooperates with the outer optimizer. The formulation of the BLPP adopted here ensures that the outer optimizer selects specific inner variables to minimize. Most BLPP approaches use this tie-cooperative formulation. It is important to point out that methods for the nonlinear BLPP assume that second order sufficiency conditions hold, hence the BLLPs solved have isolated unique inner level optima [26]. For books that include extensive studies on the BLPP, the reader is directed to [46, 51, 57].

2.2. KKT OPTIMALITY CONDITIONS

The KKT optimality conditions are equivalent to the inner optimization problem assuming: f, $\mathbf{h}$, and $\mathbf{g}$ are smooth, f and $\mathbf{g}$ are convex, $\mathbf{h}$ is linear in $\mathbf{y}$ at fixed $\mathbf{x}$ for every $\mathbf{x} \in X$, and one of the first-order constraint qualifications such as linear independence, Slater, Kuhn–Tucker or weak reverse convex condition holds in terms of $\mathbf{x}$ at a feasible point $\mathbf{y}^*$. Then, a necessary and sufficient condition for $\mathbf{y}^*$ to be an optimal solution to the inner level problem is that there exists (λ^*, μ^*) that satisfies:

$$\begin{aligned}
h_i(\mathbf{x}, \mathbf{y}^*) &= 0 \quad i \in I, \qquad \text{(KKT)} \\
\frac{\partial f(\mathbf{x}, \mathbf{y}^*)}{\partial \mathbf{y}^*} + \sum_{j=1}^{J} \lambda_j^* \frac{\partial g_j}{\partial \mathbf{y}^*} + \sum_{i=1}^{I} \mu_i^* \frac{\partial h_i}{\partial \mathbf{y}^*} &= 0 \\
g_j(\mathbf{x}, \mathbf{y}^*) + s_j^* &= 0, \quad j \in J, \\
\lambda_j^* s_j^* &= 0 \quad j \in J \\
\lambda_j^*,\ s_j^* &\geqslant 0, \quad j \in J,
\end{aligned}$$

where $\boldsymbol{\lambda}^*$, $\boldsymbol{\mu}^*$are, respectively, the KKT multiplier vectors of the inequality and equality constraints. It follows that a necessary condition for $(\mathbf{x}^*, \mathbf{y}^*, \lambda^*, \mu^*)$ to be an optimal solution of the BLPP, $(\mathbf{y}^*, \lambda^*, \mu^*)$ must satisfy the above conditions at fixed $\mathbf{x} = \mathbf{x}^*$. From this line of reasoning, the bilevel programming problem is transformed into a single level problem of the form:

$$\begin{aligned}
\min_{\mathbf{x}\,\mathbf{y}}\ & F(\mathbf{x}, \mathbf{y}) && (2)\\
& \text{s.t.}\\
& \mathbf{G}(\mathbf{x}, \mathbf{y}) \leqslant 0\\
& \mathbf{H}(\mathbf{x}, \mathbf{y}) \leqslant 0\\
& h_i(\mathbf{x}, \mathbf{y}) = 0 \quad i \in I,\\
& \frac{\partial f(\mathbf{x}, \mathbf{y})}{\partial \mathbf{y}} + \sum_{j=1}^{J} \lambda_j \frac{\partial g_j}{\partial \mathbf{y}} + \sum_{i=1}^{I} \mu_i \frac{\partial h_i}{\partial \mathbf{y}} = 0 \qquad \text{(s)}\\
& g_j(\mathbf{x}, \mathbf{y}) + s_j = 0, \quad j \in J,\\
& \lambda_j s_j = 0, \quad j \in J, \qquad \text{(cs)}\\
& \lambda_j,\ \ s_j \geqslant 0, \quad j \in J, \qquad \text{(cs)}\\
& \mathbf{x} \in X, \mathbf{y} \in Y.
\end{aligned}$$

Note that problem (2) is nonconvex due to the stationarity conditions (s) and the complementarity conditions (cs). Hence, the resulting single level formulation is nonlinear and nonconvex even if the original bilevel problem is linear due to the complementarity conditions. For the linear case, the complementarity conditions are the only nonlinearities in the single level transformed problem. Different approaches to tackle this difficulty include penalty function [10, 66], branch and bound [14, 15], global optimization [64] and reverse convex programming [59, 60] methods.

For the convex form of (1) solution methods in the literature generally require the following conditions at fixed $\mathbf{x}$ [26, 30, 31]: (a) f, $\mathbf{g}$, $\mathbf{h}$ are continuous and twice differentiable functions in $(\mathbf{x}, \mathbf{y})$; (b) the linear independence condition holds at $\mathbf{y} \in Y$, such that the gradients of the inner problem equality and active inequality constraints, $\nabla_x g_j(\mathbf{x}, \mathbf{y})\ \ \forall j \in J_A,\ \ \nabla_x h_i(\mathbf{x}, \mathbf{y})\ \ \forall i \in I$, are independent; (c) strict complementarity condition holds at $\mathbf{y} \in Y$; and (d) the second order sufficiency condition holds at $\mathbf{y} \in Y$.

Under the assumptions (a)–(d) on the functions in (1), the inducible region, IR, is continuous [57]. Assumptions (b) and (d) assure that the global optimum is also unique. Furthermore, the KKT optimality conditions are necessary and sufficient for locating the global optimum of the inner optimization problem when convex. However, the complementarity and the stationarity conditions introduce nonconvexities. Shimizu et al. (1997) suggest handling the complementarity conditions by an implicit enumeration such as in [15]. Still, the stationarity conditions may be highly nonlinear. Branch and bound [12, 28], descent [62] and global optimization [64] methods have been developed for the solution of the linear-quadratic BLPP. In all these methods, additional requirements for convex F and $\mathbf{G}$, quadratic f, and affine **h** and **g** are assumed such that the inner problem constraints form a convex polyhedron. A difference of convex functions programming based global optimization approach is presented for restricted types of problems with quadratic inner objective functions in [9].

For problems that involve convex F, f, $\mathbf{G}$, $\mathbf{g}$ and linear $\mathbf{H}, \mathbf{h}$, existing approaches, namely, descent [55] penalty function [6, 40, 49], and global optimization [8] proposed in the literature still assume that (a)–(d) hold, as well as the approaches for the solution of the general BLPP. However, for the general nonlinear case, the convexity assumption of the inner problem is usually relaxed to include nonconvex cases as long as KKT necessary conditions are met. Optimization methods developed for the general nonlinear case, include the relaxation and active set strategy techniques [26]. Note that the KKT conditions can no longer guarantee global optimality of the inner problem for fixed **x**. This means that, even if the transformed problem is solved by a global optimization approach, global optimality of the transformed single level problem can not be guaranteed. Hence, methods for the solution of the general nonlinear BLPP that are based on the KKT optimality conditions are bound to be local. The following section presents the main concepts that we have used in order to overcome the limitations of KKT-type methods.

3. Conceptual Framework

To assure that KKT optimality conditions are both necessary and sufficient for obtaining the global optimum of the inner problem, the functions f and **g** must be convex and **h** must be linear at fixed **x**.

Condition 1: If for fixed **x**, assumptions (a)–(d) hold, f and **g** are convex and **h** are linear in y, then the KKT optimality conditions are necessary and sufficient for obtaining the global optimum of the inner problem [26, 32].

If *condition 1* does not hold, then KKT conditions are only necessary. By replacing the nonconvex inner problem with its KKT optimality conditions, and solving the resulting single level problem to local optimality, an upper bound on the global optimum of the BLPP is obtained, provided that the linear independence condition holds.

3.1. UNDERESTIMATION FOR THE BLPP

A lower bound to the global optimum of the BLPP can be found as follows: the feasible region, Ω can be enlarged in such a way that the infeasible points within the convex hull are included into the feasible set. This can be done by utilizing the basic principles of the deterministic global optimization algorithm (see [32]), $\boldsymbol{\alpha}$**BB** [2, 4, 5, 11] to underestimate the nonconvex functions over the (**x**, **y**) domain.

For the nonlinear functions, valid underestimators are generated by the decomposition of each nonlinear function into a sum of terms belonging to one of several categories: linear, bilinear, trilinear, fractional, fractional trilinear, convex, univariate concave, product of univariate concave or general nonconvex,

$$
\begin{aligned}
f(\mathbf{y}) = lt(\mathbf{y}) + ct(\mathbf{y}) + \sum_{i=1}^{bt} b_i y_{B_{i,2}} y_{B_{i,2}} + \sum_{i=1}^{tt} t_1 y_{T_{i,1}} y_{T_{i,2}} y_{T_{i,3}} \\
+ \sum_{i=1}^{ft} f_i \frac{y_{f_{i,1}}}{y_{f_{i,2}}} + \sum_{i=1}^{ftt} \frac{y_{ft_{i,1}} y_{ft_{i,2}}}{y_{ft_{i,3}}} + \sum_{i=1}^{ut} ut_i(y_i) + \sum_{i=1}^{nt} nt_i(\mathbf{y}), \text{ at } x \in X.
\end{aligned}
$$

After the terms are identified, a different convex underestimator is constructed for each class of term, and a lower bounding function is obtained.

A *Bilinear Term* $(y_1 y_2)$ with $y_1 \in [y_1^L,\ y_1^U]$ and $y_2 \in [y_2^L,\ y_2^U]$ can be underestimated by introducing a variable ω that replaces every occurrence of $(y_1 y_2)$ in the problem and satisfies the relationship [2, 7]:

$$
\begin{aligned}
y_1^L y_2 + y_2^L y_1 - y_1^L y_2^L - \omega &\leqslant 0 \\
y_1^U y_2 + y_2^U y_1 - y_1^U y_2^U - \omega &\leqslant 0 \\
-y_1^U y_2 - y_2^L y_1 + y_1^U y_2^L + \omega &\leqslant 0 \\
-y_1^L y_2 - y_2^U y_1 + y_1^L y_2^U + \omega &\leqslant 0,
\end{aligned}
$$

which defines its convex envelope.

A *Univariate Concave Term*, $ut(y)$ over $[y^L,\ y^U]$, is underestimated by a linear function of y [2]:

$$
ut(y^L) + \frac{ut(y^U) - ut(y^L)}{y^U - y^L}(y - y^L).
$$

Products of Univariate Terms $(f(y_1)g(y_2))$ are underestimated using a generalization of the method for bilinear terms (see [32, 48]). *Trilinear* $(y_1 y_2 y_3)$, *Fractional* $(\frac{y_1}{y_2})$, and *Fractional Trilinear Terms* are replaced by new variables subject to sets of constraints in a way similar to the underestimation of bilinear terms (see [32, 48]).

All other *General Nonconvex Terms* for which customized underestimators do not exist are underestimated as proposed in [32]. A function $f(y)$ is underestimated in $y \in [y^L, y^U]$ by the function $\mathcal{L}(\mathbf{y})$ defined as:

$$
\mathcal{L}(\mathbf{y}) = f(\mathbf{y}) + \sum_{i=1}^{n} \alpha_i (y_i^L - y_i)(y_i^U - y_i),
$$

where α_i's are positive scalars such that $H_f(\mathbf{y}) + 2diag(\alpha_i)$ is positive semi-definite $\forall\ \mathbf{y} \in [\mathbf{y}^L, \mathbf{y}^U]$, where $H_f(\mathbf{y})$ is the Hessian matrix of the general nonconvex term

[32]. The corresponding underestimating function $\mathcal{L}(\mathbf{y})$ is convex:

$$\begin{aligned}
\mathcal{L}(\mathbf{y}) = lt(\mathbf{y}) + ct(\mathbf{y}) + \sum_{i=1}^{bt} b_i \omega_{B_i} \\
&+ \sum_{i=1}^{tt} t_i \omega_{T_i} + \sum_{i=1}^{ft} f_i \omega_{F_i} + \sum_{i=1}^{ftt} ft_i \omega_{FT_i} \\
&+ \sum_{i=1}^{ut} (ut_i(y_i^L) + \frac{ut_i(y_i^U) - ut_i(y_i^L)}{y_i^U - y_i^L}(y - y_i^L)) \\
&+ \sum_{i=1}^{nt} (nt_i(\mathbf{y}) + \sum_{j=1}^{n} \alpha_{ij}(y_j^L - y_j)(x_j^U - x_j))
\end{aligned}$$

where ω includes ω_{B_i}, ω_{T_i}, ω_{F_i}, and ω_{FT_i} variables. See [32] for rigorous calculation methods of the $\boldsymbol{\alpha}$'s.

3.1.1. *Equality Constraints of the Inner Problem*

As described in the previous section, the equality constraints of the inner problem must be linear for the KKT conditions to be necessary and sufficient. The bilinear, trilinear, fractional and fractional trilinear terms are replaced by new variables that are defined by the introduction of additional convex inequality constraints. Thus, if the equality constraint involves only these classes of terms, the resulting problem is linear. If this is not the case, and convex, univariate concave, or general nonconvex terms exist, the constraint is simply eliminated by a transformation into two inequality constraints:

$$\begin{aligned}
\mathbf{h}(\mathbf{x}, \mathbf{y}) &\leqslant \mathbf{0} \\
-\mathbf{h}(\mathbf{x}, \mathbf{y}) &\leqslant \mathbf{0},
\end{aligned}$$

which are added to the set of inequality constraints. The resulting inner problem includes the set of linear $\mathbf{h}$ and nonconvex $\boldsymbol{f}$ and $\mathbf{g}$. Note that now $\mathbf{g}$ also includes the nonlinear $\mathbf{h}$ that are written as two inequality constraints.

3.1.2. *Inequality Constraints of the Inner Problem*

Based on the underestimation of every term, a convex underestimator for any given twice-differentiable function can be obtained through the decomposition approach (See Fig.1). Let $\boldsymbol{f}^c$, $\mathbf{g}^c$ denote the convexified $\boldsymbol{f}$ and $\mathbf{g}$, respectively. Then, assuming the linear independence condition holds for the set of constraints at $\mathbf{y}^*$, there exists $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ such that the inner problem can be replaced with its equivalent KKT optimality conditions that are both necessary and sufficient for the convexified problem. In other words, for fixed $\mathbf{x}$, $\mathbf{y}^*$ is the global optima of the inner convexified problem.

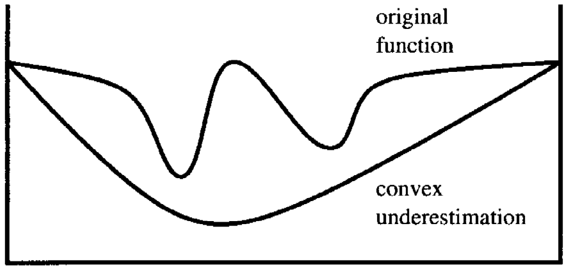

Figure 1. Convex underestimation.

After the KKT transformation of the convexified inner problem, the resulting single level problem is still nonlinear and nonconvex due to the complementarity (**cs**) and stationarity (**s**) conditions, and nonconvexities in **G**, **H**, F. The complementarity conditions are transformed into a set of MI(N)LP equations as described below. The resulting MI(N)LP problem is solved to global optimality by using one of the deterministic global optimization algorithms **SMIN-αBB** or **GMIN-αBB** [1, 3] as described below. The solution is a lower bound on the original BLPP minimum.

3.2. LINEAR INDEPENDENCE

Note that in order to replace the convexified inner problem with its equivalent KKT optimality conditions, a first order constraint qualification such as the linear independence condition of the inner problem constraints at the optima must be satisfied. Otherwise, the transformed single level problem may be infeasible or it can not be guaranteed that it is a lower bound. A simple linear independence check can be made by testing whether the best $(\mathbf{x}^*, \mathbf{y}^*)$ values obtained from the solution of the original nonconvex upper bounding problem result in linearly independent active constraints in the convexified problem.

3.3. COMPLEMENTARITY CONDITIONS – ACTIVE SETS

The complementarity condition constraints are one of the major difficulties in solving the transformed single level problem. They involve discrete decisions on the choice of the set of inner problem active constraints. The active set changes when at least one inequality function and its multiplier are equal to zero. With the change in the active set of constraints, the feasible space of the inner problem, at fixed **x**, also changes. Furthermore, the overall feasible space changes, as it is composed of different regions that correspond to different active sets. To overcome this difficulty, the ideas of active set strategy [35] can be employed. In this case a binary variable, Y_j, is introduced associated with each inequality constraint, $j \in J$, depending whether it is active or inactive. Hence, the complementarity conditions can be reformulated as follows:

$$
\begin{aligned}
&\lambda_j^c - U\,Y_j \leqslant 0 \quad j \in J,\\
&s_j^c - U\,(1 - Y_j) \leqslant 0 \quad j \in J,\\
&\lambda_j^c,\ s_j^c \geqslant 0, \quad j \in J\\
&\mathbf{x} \in X, \mathbf{y} \in Y, Y_j \in \{0, 1\},
\end{aligned}
$$

where U is an upper bound for the slack variables, s_j^c. This way, the inner problem feasible regions belonging to different active sets can be visited simultaneously as the outer problem decision vector $\mathbf{x}$ changes.

3.4. GLOBAL OPTIMIZATION OF NONLINEAR MIXED INTEGER PROBLEMS

The Special structure Mixed Integer Nonlinear αBB, **SMIN-$\boldsymbol{\alpha}$BB** [1, 3] algorithm is a deterministic global optimization method based on a branch-and-bound framework. Epsilon convergence is guaranteed for convergence to the global minimum of problems that involve separable MINLP functions that are twice-differentiable in continuous variables. A valid upper bound on the global solution is obtained by solving the nonconvex MINLP to local optimality. A lower bound is determined by solving a valid convex MINLP underestimation of the original problem. Convergence is obtained by the refinement of the feasible space into smaller regions in which convex underestimators are generated [1, 3]. The General structure Mixed Integer Nonlinear αBB, **GMIN-$\boldsymbol{\alpha}$BB** [1, 3] algorithm, on the other hand, is also based on a branch and bound framework, but is further applicable to problems without the restriction of the separability of the integer variables.

3.5. BRANCHING AND BOUNDING

After upper and lower bounds on the original BLPP problem are obtained, the initial region of ($\mathbf{x}$, $\mathbf{y}$) is partitioned into smaller regions, in the following way: Tighter lower bounds to the problem can be obtained by dividing the initial feasible region into two subregions by using one of the branching rules that are developed within the deterministic global optimization algorithm, **$\boldsymbol{\alpha}$BB** [2, 4, 5, 11]. For example, the initial feasible region, as determined by the variable bounds, can be subdivided into two subregions by halving along the longest side (bisection). After branching, minimization is performed in each subregion. The smallest minimum for all subregions of the original feasible region is the overall lower bound. At the next iteration, only the subrectangle responsible for the overall minimum is further bisected. Hence, a nondecreasing sequence of lower bounds is produced. A non-increasing sequence of upper bounds is created by locally solving the original nonconvex bilevel problem in each subregion, where the upper bound is the minimum of all upper bounds calculated in previous iterations. The upper and lower bounds bracket the global minimum. The branch and bound framework also includes a fathoming step, where any subregion with a lower bound higher than the current upper bound is removed from further consideration.

The steps of the proposed global optimization framework for the nonlinear bilevel optimization problem are presented in the following section.

4. Global Optimization Algorithm

Step 1: Set the lower bound, $z^{LO} = -\infty$, upper bound, $z^{UP} = \infty$, iteration counter $k = 1$ and select a convergence tolerance ϵ.

Step 2: Substitute the original inner optimization problem:

$$\begin{aligned}
&\min_{\mathbf{y}} \ f(\mathbf{x}, \mathbf{y}) \\
&\quad \text{s.t.} \\
&\quad h(\mathbf{x}, \mathbf{y}) = 0 \\
&\quad g(\mathbf{x}, \mathbf{y}) \leqslant 0,
\end{aligned}$$

with its KKT optimality conditions, (KKT), employ the active set strategy for the complementarity conditions of problem (2), and construct the following single stage MI(N)LP optimization problem:

$$\begin{aligned}
z = &\min_{\mathbf{x}\,\mathbf{y}} \ F(\mathbf{x}, \mathbf{y}) \\
&\quad \text{s.t.} \\
&\quad G(\mathbf{x}, \mathbf{y}) \leqslant 0 \\
&\quad H(\mathbf{x}, \mathbf{y}) = 0 \\
&\quad h_i(\mathbf{x}, \mathbf{y}) = 0 \quad i \in I, \\
&\quad \frac{\partial f(\mathbf{x}, \mathbf{y})}{\partial \mathbf{y}} + \sum_{j=1}^{J} \lambda_j \frac{\partial g_j(\mathbf{x}, \mathbf{y})}{\partial \mathbf{y}} + \sum_{i=1}^{I} \mu_i \frac{\partial h_i(\mathbf{x}, \mathbf{y})}{\partial \mathbf{y}} = 0, \\
&\quad g_j(\mathbf{x}, \mathbf{y}) + s_j \leqslant 0 \quad j \in J, \\
&\quad - g_j(\mathbf{x}, \mathbf{y}) - s_j \leqslant 0 \quad j \in J, \\
&\quad \lambda_j - U\, Y_j \leqslant 0 \quad j \in J, \\
&\quad s_j - U\,(1 - Y_j) \leqslant 0 \quad j \in J, \\
&\quad \lambda_j, \ \ s_j \geqslant 0, \quad j \in J, \\
&\quad \mathbf{x} \in \mathbf{x}, \mathbf{y} \in Y, \mathbf{y} \in \{0, 1\}.
\end{aligned}$$

Step 3: Solve the resulting problem by using a local MINLP optimizer, such as MINOPT [56] or DICOPT [63], which will yield the upper bound, z^{UP}.

Step 4: Transform the nonlinear equality constraints $\mathbf{h}(\mathbf{x}, \mathbf{y})$ into two inequality constraints $\mathbf{h}(\mathbf{x}, \mathbf{y}) \leqslant 0$ and $-\mathbf{h}(\mathbf{x}, \mathbf{y}) \leqslant 0$. The transformed constraints are now in the set $\mathbf{g}(\mathbf{x}, \mathbf{y})$. The remaining equality constraints are linear. Denote them as $\mathbf{h}^l(\mathbf{x}, \mathbf{y})$. Then, develop convex underestimators of the nonlinear terms in the functions $\mathbf{g}(\mathbf{x}, \mathbf{y})$ and $f(\mathbf{x}, \mathbf{y})$, using the basic principles of the deterministic global optimization algorithm, $\boldsymbol{\alpha}$**BB** [2, 4, 5, 11]. Denote the underestimated functions as $f^c(\mathbf{x}, \mathbf{y})$, $\mathbf{g}^c(\mathbf{x}, \mathbf{y})$.

Step 5: Establish tight upper and lower bounds on inner variables that participate in nonconvex terms that are underestimated, by solving:

$$
\begin{aligned}
y_n^L/y_n^U = \min_{\mathbf{x},\mathbf{y}} \; & y_n/-y_n \\
& h_i(\mathbf{x}, \mathbf{y}) = 0, \quad i = 1, \ldots, I \\
& g_j(\mathbf{x}, \mathbf{y}) \leqslant 0, \quad j = 1, \ldots, J,
\end{aligned}
$$

for $n = 1, \ldots, N$. and add the simple bounds thus obtained to the set of constraints [47].

Step 6: Substitute the KKT optimality conditions that are necessary and sufficient for the solution of the convexified problem:

$$
\begin{aligned}
z = \min_{\mathbf{x}\,\mathbf{y}} \; & F(\mathbf{x}, \mathbf{y}) \\
& \text{s.t.} \\
& G(\mathbf{x}, \mathbf{y}) \leqslant 0 \\
& H(\mathbf{x}, \mathbf{y}) = 0 \\
& h_i^l(\mathbf{x}, \mathbf{y}) = 0 \quad i \in I, \\
& \frac{\partial f^c(\mathbf{x}, \mathbf{y})}{\partial \mathbf{y}} + \sum_{j=1}^{J} \lambda_j^c \frac{\partial g_j^c(\mathbf{x}, \mathbf{y})}{\partial \mathbf{y}} + \sum_{i=1}^{I} \mu_i^c \frac{\partial h_i^l(\mathbf{x}, \mathbf{y})}{\partial \mathbf{y}} \leqslant 0, \\
& -\frac{\partial f^c(\mathbf{x}, \mathbf{y})}{\partial \mathbf{y}} - \sum_{j=1}^{J} \lambda_j^c \frac{\partial g_j^c(\mathbf{x}, \mathbf{y})}{\partial \mathbf{y}} - \sum_{i=1}^{I} \mu_i^c \frac{\partial h_i^l(\mathbf{x}, \mathbf{y})}{\partial \mathbf{y}} \leqslant 0, \\
& g_j^c(\mathbf{x}, \mathbf{y}) + s_j^c \leqslant 0 \quad j \in J, \\
& -g_j(\mathbf{x}, \mathbf{y}) - s_j^c \leqslant 0 \quad j \in J, \\
& \lambda_j^c s_j^c = 0 \quad j \in J, \\
& \lambda_j^c, \; s_j^c \geqslant 0, \quad j \in J, \\
& \mathbf{x} \in \mathbf{x}, \mathbf{y} \in Y.
\end{aligned}
$$

Notice that when the slack variable s_j^c is added to the convexified inequality constraint, g_j^c, an equality constraint is obtained of the form: $g_j^c + s_j^c = 0$, which is nonconvex when g_j^c is nonlinear. When applying a global optimization algorithm to the above single level optimization problem, this constraint is rewritten as two inequality constraints, of the form: $g_j^c + s_j^c \leqslant 0$ and $-g_j^c - s_j^c \leqslant 0$. Since the term g_j^c is convex, the term $-g_j^c$ in the additional inequality constraint is inherently concave, and thus needs to be underestimated. However, underestimation of this term will result in an overestimation of the term g_j^c. Thus, in order to obtain a valid lower bound to the original BLPP, the set of equations should be of the form: $g_j^c + s_j^c \leqslant 0$, and $-g_j - s_j^c \leqslant 0$, resulting in the correct underestimation of the negative of the original term, $-g_j$.

Step 7: Convert the complementarity conditions into an equivalent set of constraints involving discrete variables [35] as described in the previous section. The resulting MI(N)LP problem becomes:

$$
\begin{aligned}
z = \min_{\mathbf{x}\,\mathbf{y}} \quad & F(\mathbf{x}, \mathbf{y}) \\
& \text{s.t.} \\
& G(\mathbf{x}, \mathbf{y}) \leqslant 0 \\
& H(\mathbf{x}, \mathbf{y}) = 0 \\
& h_i^l(\mathbf{x}, \mathbf{y}) = 0 \quad i \in I, \\
& \frac{\partial f^c(\mathbf{x}, \mathbf{y})}{\partial \mathbf{y}} + \sum_{j=1}^{J} \lambda_j^c \frac{\partial g_j^c(\mathbf{x}, \mathbf{y})}{\partial \mathbf{y}} + \sum_{i=1}^{I} \mu_i^c \frac{\partial h_i^l(\mathbf{x}, \mathbf{y})}{\partial \mathbf{y}} \leqslant 0, \\
& -\frac{\partial f^c(\mathbf{x}, \mathbf{y})}{\partial \mathbf{y}} - \sum_{j=1}^{J} \lambda_j^c \frac{\partial g_j^c(\mathbf{x}, \mathbf{y})}{\partial \mathbf{y}} - \sum_{i=1}^{I} \mu_i^c \frac{\partial h_i^l(\mathbf{x}, \mathbf{y})}{\partial \mathbf{y}} \leqslant 0, \\
& g_j^c(\mathbf{x}, \mathbf{y}) + s_j^c \leqslant 0 \quad j \in J, \\
& -g_j(\mathbf{x}, \mathbf{y}) - s_j^c \leqslant 0 \quad j \in J, \\
& \lambda_j^c - U\, Y_j^c \leqslant 0 \quad j \in J, \\
& s_j^c - U\,(1 - Y_j^c) \leqslant 0 \quad j \in J, \\
& \lambda_j^c,\ s_j^c \geqslant 0, \quad j \in J, \\
& \mathbf{x} \in \mathbf{x}, \mathbf{y} \in Y, \mathbf{y} \in \{0, 1\},
\end{aligned}
$$

where s_j^c are the slack variables associated with the inner problem convexified inequalities, λ_j^c are the associated Lagrange multipliers, Y_j^c are the binary variables involved in the active set strategy. Check if the linear independence condition is satisfied at the best upper bound value obtained. If not, terminate. If satisfied, continue to Step 8.

Step 8: Solve the resulting MINLP problem to global optimality by the **SMIN-αBB** or **GMIN-αBB** algorithms [1, 3]. If the optimum outer objective function, z^*, is higher than the current lower bound, update $z^{LO} = z^*$.

Step 9: If $z^{UP} - z^{LO} \leqslant \epsilon$, stop. The global optimum is obtained. Else, go to the following step for partitioning.

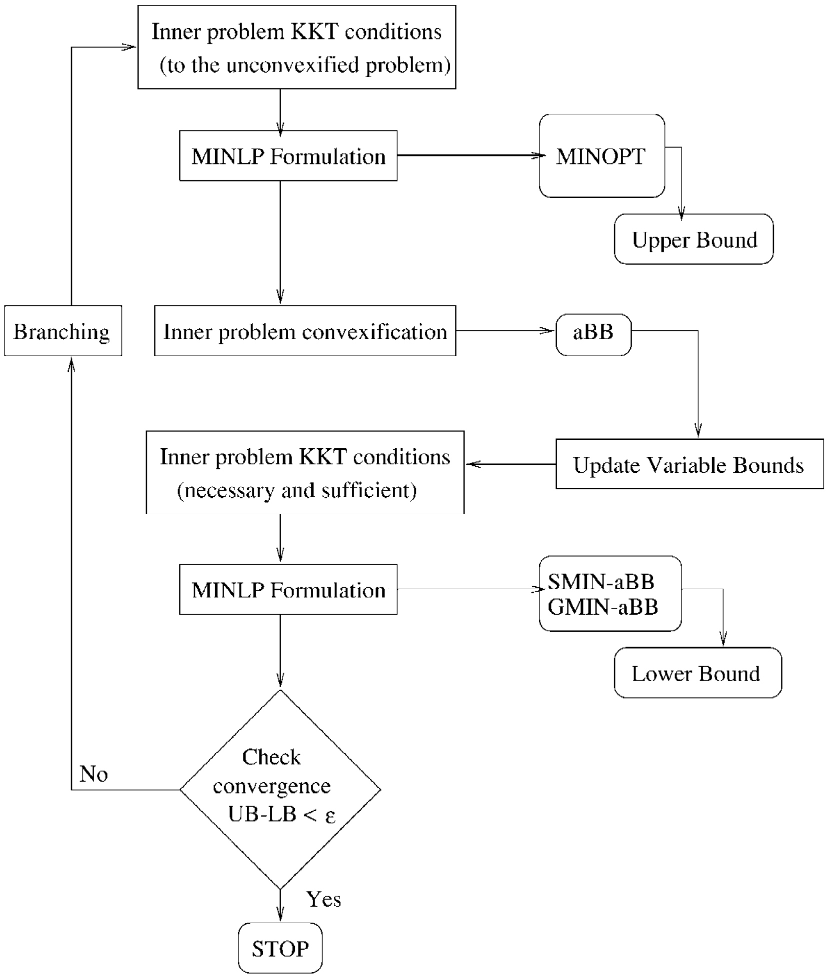

Figure 2. Algorithmic framework.

Step 10: Branch on a selected variable that participates in one of the nonlinear terms, to partition the initial domain into two subdomains to be considered at the next iteration. The branching strategy has a significant effect on the performance of the algorithm. One of the seven alternative branching strategies implemented within the $\boldsymbol{\alpha}$**BB** global optimization algorithm, as described in [2, 4]. After branching, go back to Step 2.

5. Computational Studies

EXAMPLE 1. Consider the following BLPP [57]:

$$
\begin{aligned}
\min_{x,y}\ & 16x^2 + 9y^2 \\
& \text{s.t.} \\
& -4x + y \leqslant 0, \\
& -x \leqslant 0 \\
& \min_{\mathbf{y}} (x + y - 20)^4 \\
& \qquad \text{s.t.} \\
& \qquad 4x + y - 50 \leqslant 0, \\
& \qquad -y \leqslant 0.
\end{aligned}
$$

This problem has two local optima, (7.2,12.8) and (11.25,5), where (11.25,5) is the global solution. Note that all the inner level functions are convex, and therefore no underestimation is necessary, since the inner problem KKT conditions are both necessary and sufficient. Hence the upper and lower bounding problems are the same. After replacing the inner level problem with its equivalent KKT optimality conditions, and using the active set strategy to rewrite the complementarity conditions, the resulting is an MINLP problem:

$$
\begin{aligned}
\min_{x,y}\ & 16x^2 + 9y^2 \\
& \text{s.t.} \\
& -4x + y \leqslant 0, \\
& -x \leqslant 0 \\
& 4x + y - 50 + s_1 = 0, \\
& -y + s_2 = 0 \\
& 4(x + y - 20)^3 + \lambda_1 - \lambda_2 \leqslant 0 \\
& -4(x + y - 20)^3 - \lambda_1 + \lambda_2 \leqslant 0 \\
& \lambda_j - UY_j \leqslant 0, \quad \forall j = 1, 2 \\
& s_j + UY_j \leqslant U, \quad \forall j = 1, 2 \\
& \lambda_j, s_j \leqslant 0, \quad \forall j = 1, 2,
\end{aligned}
$$

that is solved to its global optimum value by using **SMIN-$\boldsymbol{\alpha}$BB** [1, 3] in one overall iteration to: $(x, y) = (11.25, 5)$, $F^* = 2250$ in 4 **$\boldsymbol{\alpha}$BB** iterations and 2.210 CPUs on an HP J2240 using one CPU.

EXAMPLE 2. Consider the following BLPP [34]:

$$
\begin{aligned}
\min_{x,\mathbf{y}} \quad & x^3 y_1 + y_2 \\
& \text{s.t.} \\
& x \leqslant 1 \\
& -x \leqslant 0 \\
& \min_{\mathbf{y}} -y_2 \\
& \qquad x y_1 \leqslant 10 \\
& \qquad y_1^2 + x y_2 \leqslant 1 \\
& \qquad -y_2 \leqslant 0.
\end{aligned}
$$

Note that at fixed x, the inner problem is convex, so no underestimation is required for the inner problem. The KKT optimality conditions are necessary and sufficient, and the upper and lower bounding problem formulations are the same. The resulting transformed single level problem is nonlinear:

$$
\begin{aligned}
\min_{x,\mathbf{y}} \quad & x^3 y_1 + y_2 \\
& \text{s.t.} \\
& x \leqslant 1 \\
& -x \leqslant 0 \\
& x y_1 + s_1 = 10 \\
& y_1^2 + x y_2 + s_2 \leqslant 1 \\
& -y_1^2 - x y_2 - s_2 \leqslant -1 \\
& -y_2 + s_3 = 0 \\
& x\lambda_1 + 2y_1\lambda_2 \leqslant 0 \\
& -x\lambda_1 - 2y_1\lambda_2 \leqslant 0 \\
& -1 + x\lambda_2 - \lambda_3 \leqslant 0 \\
& 1 - x\lambda_2 + \lambda_3 \leqslant 0.
\end{aligned}
$$

The deterministic global optimization algorithm, **SMIN-αBB** [1, 3] is employed to locate the global optimum $(x^*, y_1^*, y_2^*) = (1, 0, -1)$, in 4 αBB iterations and 3.38 CPUs.

EXAMPLE 3. Consider the following simple fractional BLPP with linear constraints [22]:

$$
\begin{aligned}
&\min_{\mathbf{x}} -8x_1 - 4x_2 + 4y_1 - 40y_2 - 4y_3\\
&\quad \text{s.t.}\\
&\quad \min_{\mathbf{y}} \frac{1 + x_1 + x_2 + 2y_1 - y_2 + y_3}{6 + 2x_1 + y_1 + y_2 - 3y_3}\\
&\qquad \text{s.t.}\\
&\qquad -y_1 + y_2 + y_3 + y_4 = 1\\
&\qquad 2x_1 - y_1 + 2y_2 - 1/2y_3 + y_5 = 1\\
&\qquad 2x_2 + 2y_1 - y_2 - 1/2y_3 + y_6 = 1\\
&\qquad x_1, x_2, y_i \geqslant 0, \quad \forall i = 1, \ldots, 6.
\end{aligned}
$$

The problem can be put into a more tractable form by rearranging the inner objective fractional term by introducing new variables $w_1,\ w_2$ and substituting:

$$
\begin{aligned}
&w_1 = (1 + x_1 + x_2 + 2y_1 - y_2 + y_3)/w_2\\
&w_2 = 6 + 2x_1 + y_1 + y_2 - 3y_3,
\end{aligned}
$$

where the resulting problem becomes:

$$
\begin{aligned}
&\min_{\mathbf{x}} -8x_1 - 4x_2 + 4y_1 - 40y_2 - 4y_3\\
&\qquad \text{s.t.}\\
&\qquad \min_{\mathbf{y}} w_1\\
&\qquad\quad \text{s.t.}\\
&\qquad\quad -y_1 + y_2 + y_3 + y_4 = 1\\
&\qquad\quad 2x_1 - y_1 + 2y_2 - 1/2y_3 + y_5 = 1\\
&\qquad\quad 2x_2 + 2y_1 - y_2 - 1/2y_3 + y_6 = 1\\
&\qquad\quad 1 + x_1 + x_2 + 2y_1 - y_2 - y_3 - w_1 w_2 = 0\\
&\qquad\quad 6 + 2x_1 + y_1 + y_2 - 3y_3 - w_1 = 0\\
&\qquad\quad x_1, x_2, y_i \geqslant 0, \quad \forall i = 1, \ldots, 6.
\end{aligned}
$$

Note that the $w_1 w_2$ term is bilinear, hence the inner problem is a nonconvex BLPP. The bilinear term can be underestimated by introducing a variable ω that replaces every occurrence of $w_1 w_2$ in the problem and satisfies the constraints that define its convex envelope [2, 7]. The underestimating problem can be reformulated over the variable domain:

$$
\begin{aligned}
\min_{\mathbf{x}}\ & -8x_1 - 4x_2 + 4y_1 - 40y_2 - 4y_3 \\
& \text{s.t.} \\
& \min_{\mathbf{y}}\ w_1 \\
& \quad \text{s.t.} \\
& \quad -y_1 + y_2 + y_3 + y_4 = 1 \\
& \quad 2x_1 - y_1 + 2y_2 - 1/2y_3 + y_5 = 1 \\
& \quad 2x_2 + 2y_1 - y_2 - 1/2y_3 + y_6 = 1 \\
& \quad 1 + x_1 + x_2 + 2y_1 - y_2 - y_3 - w_3 = 0 \\
& \quad 6 + 2x_1 + y_1 + y_2 - 3y_3 - w_1 = 0 \\
& \quad w_1^L w_2 + w_2^L w_1 - w_1^L w_2^L - w_3 \leqslant 0 \\
& \quad w_1^U w_2 + w_2^U w_1 - w_1^U w_2^U - w_3 \leqslant 0 \\
& \quad -w_1^U w_2 - w_2^L w_1 + w_1^U w_2^L + w_3 \leqslant 0 \\
& \quad -w_1^L w_2 - w_2^U w_1 + w_1^L w_2^U + w_3 \leqslant 0 \\
& \quad x_1, x_2, y_i \geqslant 0,\ \forall i = 1, \ldots, 6.
\end{aligned}
$$

The resulting inner problem is convex. Replacing with its corresponding KKT conditions that are necessary and sufficient, and introducing a binary variable Y_j for every inner constraint j, the transformed single level problem becomes:

$$
\begin{aligned}
\min_{\mathbf{x}}\ & -8x_1 - 4x_2 + 4y_1 - 40y_2 - 4y_3 \\
& \text{s.t.} \\
& -y_1 + y_2 + y_3 + y_4 = 1 \\
& 2x_1 - y_1 + 2y_2 - 1/2y_3 + y_5 = 1 \\
& 2x_2 + 2y_1 - y_2 - 1/2y_3 + y_6 = 1 \\
& 1 + x_1 + x_2 + 2y_1 - y_2 - y_3 - w_3 = 0 \\
& 6 + 2x_1 + y_1 + y_2 - 3y_3 - w_1 = 0 \\
& w_1^L w_2 + w_2^L w_1 - w_1^L w_2^L - w_3 + s_1 = 0 \\
& w_1^U w_2 + w_2^U w_1 - w_1^U w_2^U - w_3 + s_2 = 0 \\
& -w_1^U w_2 - w_2^L w_1 + w_1^U w_2^L + w_3 + s_3 = 0 \\
& -w_1^L w_2 - w_2^U w_1 + w_1^L w_2^U + w_3 + s_4 = 0 \\
& -y_i + s_{i+4} = 0,\ \forall i = 1, \ldots, 6
\end{aligned}
$$

$$
\begin{aligned}
& -\mu_1 - \mu_2 + 2\mu_3 + 2\mu_4 + \mu_5 - \lambda_5 = 0 \\
& \mu_1 + 2\mu_2 - 2\mu_3 - \mu_4 + \mu_5 - \lambda_6 = 0 \\
& \mu_1 - 1/2\mu_2 - 1/2\mu_3 - \mu_4 - 3\mu_5 - \lambda_7 = 0 \\
& \mu_1 - \lambda_8 = 0 \\
& \mu_2 - \lambda_9 = 0 \\
& \mu_3 - \lambda_{10} = 0 \\
& 1 + w_2^L\lambda_1 + w_2^U\lambda_2 - w_2^L\lambda_3 - w_2^U\lambda^4 = 0 \\
& -\mu_4 - \lambda_1 - \lambda_2 + \lambda_3 + \lambda_4 = 0 \\
& -\mu_5 + w_1^L\lambda_1 + w_1^U\lambda_2 - w_1^U\lambda_3 - w_1^L\lambda_4 = 0 \\
& \lambda_j - UY_j \leqslant 0, \quad \forall j = 1, \dots, 10 \\
& s_j + UY_j \leqslant U, \quad \forall j = 1, \dots, 10 \\
& \lambda_j,\ s_j \leqslant 0, \quad \forall j = 1, \dots, 10 \\
& x_1, x_2, y_i \geqslant 0, \quad \forall i = 1, \dots, 6,
\end{aligned}
$$

where $\boldsymbol{\lambda}, \boldsymbol{\mu}$ are Lagrange multipliers and $\mathbf{s}$ are the slack variables. The bounds $w_1^L = 0, w_1^U = 1, w_2^L = 1.0, w_2^U = 8.0, w_3^L = 0, w_3^U = 2.75$ are estimated from the solution of:

$$
\begin{aligned}
\min_{\mathbf{x},\mathbf{y}} \ & w_k,\ -w_k \\
& \text{s.t.} \\
& -y_1 + y_2 + y_3 + y_4 = 1 \\
& 2x_1 - y_1 + 2y_2 - 1/2y_3 + y_5 = 1 \\
& 2x_2 + 2y_1 - y_2 - 1/2y_3 + y_6 = 1 \\
& 1 + x_1 + x_2 + 2y_1 - y_2 - y_3 - w_3 = 0 \\
& 6 + 2x_1 + y_1 + y_2 - 3y_3 - w_1 = 0, \quad \forall k = 1, \dots, 3,
\end{aligned}
$$

taking 1 αBB iteration in around 0.030 CPUs per bound. The MILP problem is solved to global optimality at $(x_1^*,\ x_2^*,\ y_1^*,\ y_2^*,\ y_3^*,\ y_4^*,\ y_5^*,\ y_6^*,\ w_1,\ w_2,\ w_3) = (0.0,\ 0.9,\ 0.0,\ 0.6,\ 0.4,\ 0.0,\ 0.0,\ 0.0,\ 0.1125,\ 5.4,\ 0.9)$, $F^{LB} = -29.2$, using SMIN-αBB [3, 1] in 1 iteration and 0.340 CPUs. Note that the problem is linear and thus can also be solved using a linear optimization package.

The upper bounding problem is obtained by replacing the inner level problem with its KKT optimality conditions without underestimation:

$$
\begin{aligned}
\min_{\mathbf{x}}\ & -8x_1 - 4x_2 + 4y_1 - 40y_2 - 4y_3 \\
& \text{s.t.} \\
& -y_1 + y_2 + y_3 + y_4 = 1 \\
& 2x_1 - y_1 + 2y_2 - 1/2y_3 + y_5 = 1 \\
& 2x_2 + 2y_1 - y_2 - 1/2y_3 + y_6 = 1 \\
& 1 + x_1 + x_2 + 2y_1 - y_2 - y_3 - w_1w_2 = 0 \\
& 6 + 2x_1 + y_1 + y_2 - 3y_3 - w_1 = 0 \\
& -y_i + s_i = 0, \quad \forall i = 1, \ldots, 6 \\
& -\mu_1 - \mu_2 + 2\mu_3 + 2\mu_4 + \mu_5 - \lambda_1 = 0 \\
& \mu_1 + 2\mu_2 - 2\mu_3 - \mu_4 + \mu_5 - \lambda_2 = 0 \\
& \mu_1 - 1/2\mu_2 - 1/2\mu_3 - \mu_4 - 3\mu_5 - \lambda_3 = 0 \\
& \mu_1 - \lambda_4 = 0 \\
& \mu_2 - \lambda_5 = 0 \\
& \mu_3 - \lambda_6 = 0 \\
& 1 - w_2\mu_4 = 0 \\
& -w_1\mu_4 - \mu_5 = 0 \\
& \lambda_j - UY_j \leqslant 0, \quad \forall j = 1, \ldots, 6 \\
& s_j + UY_j \leqslant U, \quad \forall j = 1, \ldots, 6 \\
& \lambda_j,\ s_j \leqslant 0, \quad \forall j = 1, \ldots, 6 \\
& x_1, x_2, y_i \geqslant 0, \quad \forall i = 1, \ldots, 6.
\end{aligned}
$$

Solving the MINLP to local optimality using MINOPT [56] resulted in $F^{UB} = -29.2 = F^{LB} = F^*$, and the algorithm terminates.

EXAMPLE 4. Consider the following problem with constraints only on the outer problem [54]:

$$
\begin{aligned}
\min_{\mathbf{x}}\ & (x-3)^2 + (y-2)^2 \\
& \text{s.t.} \\
& -2x + y - 1 \leqslant 0 \\
& x - 2y + 2 \leqslant 0 \\
& x + 2y - 14 \leqslant 0 \\
& 0 \leqslant x \leqslant 8 \\
& \min_{\mathbf{y}} (y-5)^2
\end{aligned}
$$

Note that the inner problem is quadratic in y, thus no underestimation is needed for the inner problem. Furthermore, since there are no inner constraints, after the KKT transformation, the resulting problem is a quadratic optimization problem, without integer variables, and can thus be solved to global optimality by even using a local optimization method. The problem is optimal at (3,5).

EXAMPLE 5. Consider the following example problem, with the inner problem taken from [33] at $t_3^* = 0.565$:

$$
\begin{aligned}
&\min_{t_2}\ t_2\\
&\quad \text{s.t.}\\
&\quad \min_{t_1,t_4}\ -t_1 + 0.5864 t_1^{0.67}\\
&\qquad \text{s.t.}\\
&\qquad 0.0332333 t_4^{-1} + 2 t_2^{-0.71} t_4^{-1} + 0.0332333 t_2^{-1.3} \leqslant 1\\
&\qquad t_i \leqslant 10, \quad \forall i \in I = \{1,2,4\}\\
&\qquad -t_i \leqslant 0.1 \quad \forall i \in I = \{1,2,4\}.
\end{aligned}
$$

The problem constraints are nonlinear and nonconvex. However, all the nonlinearities are in terms of powers and multiplications. This special structure of the problem can be exploited by redefining the terms as: $t_1 = \exp(t_1')$, $t_2 = \exp(t_2')$, $t_4 = \exp(t_4')$ and substituting the exponential terms into the formulation:

$$
\begin{aligned}
&\min_{t_2'}\ \exp(t_2')\\
&\quad \text{s.t.}\\
&\quad \min_{t_1',t_4'}\ -\exp(t_1') + 0.5864 \exp(0.67 t_1')\\
&\qquad \text{s.t.}\\
&\qquad 0.0332333 \exp(t_4') + 0.1 \exp(t_1') \leqslant 1\\
&\qquad 4 \exp(t_2' - t_4') + 2 \exp(-0.71 t_2' - t_4') + 0.0332333 \exp(-1.3 t_2') \leqslant 1\\
&\qquad t_i' \leqslant \ln(10), \quad \forall i \in I = \{1,2,4\}\\
&\qquad -t_i' \leqslant -\ln(0.1) \quad \forall i \in I = \{1,2,4\}.
\end{aligned}
$$

Note that the resulting BLPP involves nonlinearities in the exponentials, that are either convex or concave. For the lower bounding problem, the univariate concave term $-\exp(t_1')$ is underestimated as: $-1.1052 - 2224.784(t_1' - 0.1)$. After the KKT transformation, and active set strategy, the lower bounding problem becomes:

$$
\begin{aligned}
&\min_{t_2'} \ \exp(t_2') \\
&\quad \text{s.t.} \\
&\quad 0.0332333 \exp(t_4') + 0.1 \exp(t_1') + s_1 \leqslant 1 \\
&\quad -0.0332333 \exp(t_4') - 0.1 \exp(t_1') - s_1 \leqslant -1 \\
&\quad 4 \exp(x) + 2 \exp(y) + 0.0332333 \exp(-1.3 t_2) + s_2 \leqslant 1 \\
&\quad -\exp(x) - 2 \exp(y) - 0.0332333 \exp(-1.3 t_2) - s_2 \leqslant -1 \\
&\quad t_1' + s_3 = \ln(10) \\
&\quad -t1' + s_4 = -\ln(0.1) \\
&\quad t_4' + s_5 = \ln(10) \\
&\quad -t_4' + s_6 = -\ln(0.1) \\
&\quad -2224.784 + 0.1 \exp(t_1')\lambda_1 + \lambda_3 - \lambda_4 = 0 \\
&\quad 0.0332333 \exp(t_4')\lambda_1 - 4 \exp(t_2' - t_4')\lambda_2 - 2 \exp(-0.71 t_2' - t_4')\lambda_2 \\
&\qquad\quad + \lambda_5 - \lambda_6 = 0 \\
&\quad x = t_2' - t_4' \\
&\quad y = -0.71 t_2' - t_4' \\
&\quad t_2' \leqslant \ln(10) \\
&\quad -t_2' \leqslant -\ln(0.1).
\end{aligned}
$$

The resulting MINLP is solved to global optimality by **SMIN-αBB** [1, 3] with the objective function value of -2.182606, at $(t_1', t_2', t_4') = (1.90, 0.78, 2.30)$ in 1 iteration and 0.110 CPUs. The upper bounding problem is formulated without underestimation:

$$
\begin{aligned}
&\min_{t_2'} \ \exp(t_2') \\
&\quad \text{s.t.} \\
&\quad 0.0332333 \exp(t_4') + 0.1 \exp(t_1') + s_1 = 1 \\
&\quad 4 \exp(t_2' - t_4') + 2 \exp(-0.71 t_2' - t_4') + 0.0332333 \exp(-1.3 t_2') + s_2 = 1 \\
&\quad t_1' + s_3 = \ln(10) \\
&\quad -t_1' + s_4 = -\ln(0.1) \\
&\quad t_4' + s_5 = \ln(10) \\
&\quad -t_4' + s_6 = -\ln(0.1) \\
&\quad -\exp(t_1') + 0.5864 * 0.67 \exp(0.67 t_1') + 0.1 \exp(t - 1')\lambda_1 + \lambda_3 - \lambda_4 = 0 \\
&\quad -2 \exp(-0.71 t_2' - t_4')\lambda_3 + \lambda_5 - \lambda_6 = 0 \\
&\quad t2' \leqslant \ln(10) \\
&\quad -t2' \leqslant -\ln(0.1).
\end{aligned}
$$

Solving with MINOPT [56], the upper bound of the objective function -2.182606 is obtained in 0.47 CPUs, that is equal to the lower bound, and thus the optimal solution.

6. Parameter Estimation Problems

In many science and engineering areas, estimation of parameters is a crucial step for the development of mathematical models that can accurately predict a physical phenomena. The parameters are determined from the available experimental data using statistical methods, and, in general, the models are complex and nonlinear. Statistical methods treat the experimental data measurements (the independent variables) as free of error, and only consider the error in adjustable parameters (the dependent variables). However, the independent variables contain errors associated with measurements as well. A statistical method extensively used in literature for parameter estimation where error in all the variables is treated instead of only the dependent variables is the maximum likelihood estimation method. The systems considered are described by an algebraic set of equations of the form:

$$\mathbf{f}(\boldsymbol{\theta}, \mathbf{z}) = 0,$$

where $\boldsymbol{\theta}$ is the vector of p unknown parameters, $\mathbf{z}$ is the vector of n measurement variables and $\mathbf{f}$ is the the system of l algebraic functions. The measured variables are the sum of the unknown true values $\boldsymbol{\zeta}_m$ and the additive error, $\mathbf{e}_m$ at the data point m:

$$\mathbf{z}_m = \boldsymbol{\zeta}_m + \mathbf{e}_m.$$

Obviously, $\mathbf{e}_m = \mathbf{z}_m - \boldsymbol{\zeta}_m$. Assuming that the error is normally distributed with zero mean and the covariance matrix is known, the parameters $\boldsymbol{\theta}$ are estimated from the solution of the following optimization problem:

$$\begin{aligned} \psi = \min_{\hat{\theta},\hat{z}} \quad & \sum_{m=1}^{M} (\hat{\mathbf{z}}_m - \mathbf{z}_m)^T \mathbf{V}_m^{-1} (\hat{\mathbf{z}}_m - \mathbf{z}_m) \\ & \text{s.t.} \\ & f(\hat{\mathbf{z}}_m,\ \hat{\theta}) = 0,\ m = 1, \ldots, M. \end{aligned}$$

Note that the true values of the experimental data, $\boldsymbol{\zeta}_m$ are not known, however, can be approximated from the optimization as fitted data variables $\hat{\mathbf{z}}_m$. Further, assuming that the covariance matrix is same in each experiment and diagonal, the problem becomes:

$$\begin{aligned}\psi = \min_{\hat{\theta},\hat{z}} \quad & \sum_{m=1}^{M}\sum_{i=1}^{n} \frac{(\hat{z}_{m,i} - z_{m,i})^2}{\sigma_i^2} \\ \text{s.t.} \quad & \\ & \mathbf{f}(\hat{\mathbf{z}}_m, \hat{\boldsymbol{\theta}}) = 0, \quad m = 1, \ldots, M,\end{aligned}$$

where σ_i is the standard deviation of ith variable in all experiments. This is a popular formulation of the maximum likelihood approach, the Error-in-Variables, EVM model. Notice that the objective function is convex. However, since the minimization is over both the parameters and data variables, the model equations introduce nonconvexities even for simplest cases. Solution methods for this problem include simultaneous parameter estimation and data reconciliation, two-stage nonlinear EVM and nested nonlinear EVM [42] and global optimization methods [29]. The nested nonlinear EVM has a bilevel formulation, of the form:

$$\begin{aligned}\psi_1 = \min_{\hat{\theta}} \quad & \sum_{m=1}^{M}\sum_{i=1}^{n} \frac{(\hat{z}_{m,i} - z_{m,i})^2}{\sigma_i^2} \\ \text{s.t.} \quad & \\ & \psi_2 = \min_{\hat{z}} \quad \sum_{m=1}^{M}\sum_{i=1}^{n} \frac{(\hat{z}_{m,i} - z_{m,i})^2}{\sigma_i^2} \\ & \qquad \text{s.t.} \\ & \qquad\qquad f(\hat{z}_{m,i}, \hat{\theta}) = 0, \quad m = 1, \ldots, M \quad i = 1, \ldots, N.\end{aligned}$$

Employing the nested nonlinear EVM formulation, several example problems are solved below.

EXAMPLE 1: *Kowalik Problem.* Consider the model equation [29]:

$$\hat{z}_{m,1} = \frac{\theta_1 z_{m,2}^2 + z_{m,2}\theta_2\theta_1}{z_{m,2}^2 + z_{m,2}\theta_3 + \theta_4}.$$

In this problem, it is assumed that only z_1 contains error, hence $z_{m,2}$ is a constant. The resulting BLPP is as follows:

$$
\begin{aligned}
\psi_1 = \min_{\hat{\theta}} \quad & \sum_{m=1}^{11} (\hat{z}_{m,1} - z_{m,1})^2 \\
& \text{s.t.} \\
\psi_2 = \min_{\mathbf{z}} \quad & \sum_{m=1}^{11} (\hat{z}_{m,1} - z_{m,1})^2 \\
& \text{s.t.} \\
& \hat{z}_{m,1} z_{m,2}^2 + \hat{z}_{m,1} z_{m,2} \theta_3 + \hat{z}_{m,1}\theta_4 - \theta_1 z_{m,2}^2 - z_{m,2}\theta_2\theta_1 = 0,
\end{aligned}
$$

where the model equation is rearranged into a simpler form. Notice that the inner problem is convex at constant $\boldsymbol{\theta}$, hence no underestimation is needed before the inner problem is replaced with its KKT optimality conditions, and the upper and lower bounding problems have the same formulation:

$$
\begin{aligned}
& \psi_1 = \min_{\hat{\theta}, \hat{z}_{m,1}} \quad \sum_{m=1}^{11} (\hat{z}_{m,1} - z_{m,1})^2 \\
& \text{s.t.} \\
& \hat{z}_{m,1} z_{m,2}^2 + \hat{z}_{m,1} z_{m,2} \theta_3 + \hat{z}_{m,1}\theta_4 - \theta_1 z_{m,2}^2 - z_{m,2}\theta_2\theta_1 = 0 \\
& 2(\hat{z}_{m,1} - z_{m,1}) + z_{m,2}^2 \mu_1 + z_{m,2}\theta_3\mu_m + \theta_4\mu_m - \lambda_{m,1} + \lambda_{m,2} = 0 \\
& -\hat{z}_{m,1} + \hat{z}_{m,1}^L + s_{m,1} = 0, \; m = 1, \ldots, 11 \\
& \hat{z}_{m,1} - \hat{z}_{m,1}^U + s_{m,2} = 0, \; m = 1, \ldots, 11 \\
& \lambda_{m,j} - UY_{m,j} \leqslant 0, \; m = 1, \ldots, 11, \; j = 1, 2 \\
& s_{m,j} + UY_{m,j} \leqslant U, \; m = 1, \ldots, 11, \; j = 1, 2 \\
& \lambda_{m,j} \geqslant 0, s_{m,j} \geqslant 0, \; m = 1, \ldots, 11, \; j = 1, 2,
\end{aligned}
$$

where U is a large positive number, μ_m is the Lagrange multiplier of equality constraint m, $\lambda_{m,j}$ is the Lagrange multiplier of the inequality constraint (m, j), and $Y_{m,j}$ is the binary variable associated with each active constraint (m, j). The parameter bounds are $[-0.2892, 0.2893]$. Setting the absolute convergence to 10^{-4} and solving the resulting nonconvex single level MINLP optimization problem to global optimality using the **SMIN-$\boldsymbol{\alpha}$BB** [3, 1], the objective is $F^* = 3.0747\text{x}10^{-4}$ and the parameters are $(\theta_1^*, \theta_2^*, \theta_3^*, \theta_4^*)$ = (0.1928, 0.1909, 0.1231, 0.1358). The data and fitted values are presented in Table 1 in the Appendix.

In the formulation of parameter estimation problems, binary variables are introduced to define the simple bounds on the variables explicity for the KKT optimality conditions. However, this creates a sigificant increase in the size of the problem, as two binary variables are introduced for every fitted variable, for every lower and upper bound, in addition to the associated Lagrange multiplier and slack variable. Notice that when these bounds are not active, the binary variables and the Lagrange multipliers take the value of zero, and thus have no constraining effects on the inner problem. Therefore, instead of defining the simple variable bounds explicitly, the problem can be solved without these constraints defined at the inner problem, and the resulting problem solved to global optimality. If a variable is at its lower or upper bound, then the constraint that defines this boundary can be included into the explicit formulation and solved again to global optimality. The new formulation includes a binary variable for the simple bound constraint and the associated KKT condition constraints. This can result in significant decreases in run time for the solution of parameter estimation problems. For the example solved above, the problem becomes an NLP when simple bounds are excluded, and the optimal value is obtained at the first global run using $\boldsymbol{\alpha}$**BB** [2, 4, 5, 11] in 37.970 CPUs and 422 iterations. No fitted variable is at its bound at this solution, so no new variables are introduced and the iteration terminates.

EXAMPLE 2: *Linear Fit* [29]. Consider the problem of linear fitting data to a straight line with the model equation:

$$\hat{z}_{m,2} = \theta_1 + \theta_2 \, \hat{z}_{m,1}.$$

The BLPP formulation is of the form:

$$\begin{aligned}
\psi_2 = \min_{\hat{\theta}} \quad & \sum_{m=1}^{10} \sum_{i=1}^{2} (\hat{z}_{m,i} - z_{m,i})^2 \\
\text{s.t.} \quad & \\
& \psi_2 = \min_{\mathbf{z}} \quad \sum_{m=1}^{10} \sum_{i=1}^{2} (\hat{z}_{m,i} - z_{m,i})^2 \\
& \quad \text{s.t.} \\
& \quad -\hat{z}_{m,2} + \theta_1 + \theta_2 \, \hat{z}_{m,1} = 0 \\
& \quad -\hat{z}_{m,1} + \hat{z}^L_{m,1} \leqslant 0 \\
& \quad \hat{z}_{m,1} - \hat{z}^U_{m,1} \leqslant 0 \\
& \quad -\hat{z}_{m,2} + \hat{z}^L_{m,2} \leqslant 0 \\
& \quad \hat{z}_{m,2} - \hat{z}^U_{m,2} \leqslant 0.
\end{aligned}$$

Since the inner problem is convex at constant $\boldsymbol{\theta}$, no underestimation is needed before it is replaced with its KKT optimality conditions and the upper and lower bounding problems have the same formulation:

$$\psi_2 = \min_{\hat{\theta}} \sum_{m=1}^{10} \sum_{i=1}^{2} (\hat{z}_{m,i} - z_{m,i})^2$$

$$\begin{aligned}
&\text{s.t.}\\
&2(\hat{z}_{m,1} - z_{m,1}) + \theta_2 \mu_m - \lambda_{m,1} + \lambda_{m,2} = 0\\
&2(\hat{z}_{m,2} - z_{m,2}) - \mu_m - \lambda_{m,3} + \lambda_{m,4} = 0\\
&-\hat{z}_{m,2} + \theta_1 + \theta_2 \hat{z}_{m,1} = 0\\
&-\hat{z}_{m,1} + \hat{z}^L_{m,1} + s_{m,1} = 0\\
&\hat{z}_{m,1} - \hat{z}^U_{m,1} + s_{m,2} = 0\\
&-\hat{z}_{m,2} + \hat{z}^L_{m,2} + s_{m,3} = 0\\
&\hat{z}_{m,2} - \hat{z}^U_{m,2} + s_{m,4} = 0\\
&\lambda_{m,j} \geqslant 0, s_{m,j} \geqslant 0, \quad m = 1, \ldots, 11,\ j = 1, \ldots, 4\\
&\lambda_{m,j} - UY_{m,j} \leqslant 0, \quad m = 1, \ldots, 11,\ j = 1, \ldots, 4\\
&s_{m,j} + UY_{m,j} \leqslant U, \quad m = 1, \ldots, 11,\ j = 1, \ldots, 4\\
&\lambda_{m,j},\ s_{m,j} \leqslant 0, \quad m = 1, \ldots, 11,\ j = 1, \ldots, 4.
\end{aligned}$$

Solving the resulting nonconvex single level MINLP for $z_m \pm 0.5$ to global optimality using the **SMIN-αBB** [1, 3] the objective is $F^* = 0.61857$ and parameters $\theta_1 = 5.7840$ and $\theta_2 = -0.54556$, in 28 iterations and 614.500 CPUs. Excluding the explicit formulation of the simple upper and lower bounding constraints on the variables for the inner problem, and solving the resulting NLP to global optimality using **αBB** [2, 4, 5, 11], the same objective is obtained in 1.98 CPUs and 28 **αBB** iterations. Since no variable is at its upper or lower bound, no additional variables or constraints are introduced, and thus the global optimum is obtained. The data and fitted values are presented in Table 2 in the Appendix.

7. Conclusions

A global optimization algorithm for the solution of the general nonlinear bilevel programming problem that involves twice differentiable functions is presented. The approach is based on a relaxation of the feasible region and branch and bound framework. The relaxation is accomplished by an enlargement of the feasible solution space of the bilevel problem. The resulting relaxed optimization problem is solved to global optimality by using the deterministic global optimization algorithm, **αBB** [2, 4, 5, 11] or **SMIN-αBB** or **GMIN-αBB** [1, 3] when integer variables are involved. Consequently, a lower bound is obtained. An upper bound to the global minimum is obtained by transforming the original problem into a single level one without the relaxation and solving for local optimality. After upper and lower bounds are obtained to the global solution, the initial region of the problem variables is partitioned into smaller regions by using one of the branching rules that are developed within the deterministic global optimization algorithm, **αBB**. Several examples of varying features are presented to show the capability of the approach in solving various BLPP problems.

Acknowledgements

The authors gratefully acknowledge financial support from the National Science Foundation and the ExxonMobil Company.

Appendix

Table 1. Kowalik Problem

Data		Fitted Values
$z_{\mu,1}$	$1/z_{\mu,2}$	$\hat{z}_{\mu,1}$
0.1957	0.25	0.1944
0.1947	0.5	0.1928
0.1735	1	0.1824
0.1600	2	0.1489
0.0844	4	0.0928
0.0627	6	0.0624
0.0456	8	0.0457
0.0342	10	0.0355
0.0323	12	0.0288
0.0235	14	0.0241
0.0246	16	0.0207

Table 2. Linear Fit Problem

Data		Linear Fit	
$z_{m,1}$	$z_{m,2}$	$\hat{z}_{m,1}$	$\hat{z}_{m,2}$
0.0	5.9	−0.049	5.811
0.9	5.4	0.855	5.318
1.8	4.4	1.969	4.710
2.6	4.6	2.501	4.419
3.3	3.5	3.503	3.873
4.4	3.7	4.267	3.456
5.2	2.8	5.262	2.913
6.1	2.8	5.955	2.535
6.5	2.4	6.432	2.275
7.4	1.5	7.504	1.690

References

1. Adjiman, C.S., Androulakis, I.P. and Floudas, C.A. (1997), ‘Global Optimization of minlp problems in process synthesis’. *Comp. Chem. Engng.* 21: S445.
2. Adjiman, C.S., Androulakis, I.P. and Floudas, C.A. (1998b), ‘A Global Optimization Method, αBB, for General Twice-Differentiable Constrained NLPs II. Implementation and computational results’. *Comp. Chem. Engng.* 22(9): 1159–1179.
3. Adjiman, C.S., Androulakis, I.P. and Floudas, C.A. (2000), ‘Global Optimization of Mixed-Integer Nonlinear Problems’. *AIChE J.* 46(9): 1769–1797.
4. Adjiman, C.S., Dallwig, S., Floudas, C.A. and Neumaier, A. (1998a), ‘A Global Optimization Method, αBB, for General Twice-Differentiable Constrained NLPs I. Theoretical Advances’. *Comp. Chem. Engng.* 22(9): 1137–1158.

5. Adjiman, C.S. and Floudas, C.A. (1996), Rigorous Convex Underestimators for General Twice–Differentiable Problems. *Journal of Global Optimization* 9: 23.
6. Aiyoshi, E. and Shimizu, K. (1996), Hierarchical decentralized systems and its solution by a barrier method. *IEEE Trans. on Sys., Man., Cyb.* 11(6): 444.
7. Al-Khayyal, F.A. and Falk, J.E. (1983), Jointly Constrained Biconvex Programming. *Maths. Ops. Res.* 8: 273.
8. Al-Khayyal, F.A., Horst, R. and P.P. M., (1992), Global Optimization of concave functions subject to quadratic constraints: An application in nonlinear bilevel programming. *Annals of Operations Research* 34: 125.
9. Amouzegar, M.A., (1999), A global optimization method for nonlinear bilevel programming problems. *IEEE Trans. Syst., Man, Cyber.* 29(6): 771–777.
10. Anandalingam, G. and White, D.J. (1990), A solution method for the linear static Stackelberg problem using penalty functions. *IEEE Trans. Auto. Contr.* 35(10): 1170–1173.
11. Androulakis, I.P., Maranas, C.D. and Floudas, C.A. (1995), αbb: A global optimization method for general constrained nonconvex problems. *Journal of Global Optimization* 7: 337.
12. Bard, J.F. (1988), Convex two-level optimization. *Mathematical Programming* 40: 15–27.
13. Bard, J.F. (1991), Some properties of the bilevel programming problem. *Journal of Optimization Theory and Applications* 68: 371–378.
14. Bard, J.F. and Falk, J. (1982), An explicit solution to the multi-level programming problem. *Comp. Op. Res.* 9: 77–100.
15. Bard, J.F. and Moore, J.T. (1990), A branch and bound algorithm for the bilevel programming problem. *SIAM J. on scientific and statistical computing* 11: 281–282.
16. Bard, J.F., Plummer, J. and S.J. C., (1998), Determining tax credits for converting non-food crops to biofuels: an application of bilevel programming. In: Migdalas, A., Pardalos, P.M. and Värbrand, P. (eds), *Multilevel Optimization: Algorithms and Applications*, p. 23. Nonconvex Optimization and Its Applications, p. 23. Kluwer Academic Publishers, Dordrecht/Boston/London
17. Ben-Ayed, O. and Blair, C.E. (1990), Computational Difficulties of Bilevel Linear Programming. *Operations Research* 38: 556–559.
18. Ben-Ayed, O., Boyce, D.E. and Blair III, C.E. (1988), A general bilevel linear programming problem formulation of the network design problem. *Trans. Res.* 22B(4): 311.
19. Bialas, W.F. and Karwan, M.H. (1984), Two-level Linear Programming. *Management Science* 30: 1004.
20. Bracken, J. and McGill, J.T. (1973), A convex programming model for optimization problems in the constraints. *Operations Research* 21.
21. Brengel, D.D. and Seiderm, W. (1992), Coordinated design and control optimization of nonlinear processes. *Comp. Chem. Engng.* 16: 861–886.
22. Calvete, H.J. and Gale, C. (1999), The bilevel linear/linear fractional programming problem. *European Journal of Operational Research* 114: 188–197.
23. Candler, W. and Norton, N. (1977), Multi-level programming and development policy. Technical Report Working Paper No. 258, World Bank.
24. Candler, W. and Townsley, R. (1982), A linear two-level programming problem. *Comp. Op. Res.* 9(1).
25. Cassidy, R.G., Kriby, M.J.L. and Raike, W.M. (1971), Efficient distribution of resources through three levels of government. *Management Science* 17(8): B–462.
26. Clark, P.A. and Westerberg, A.W. (1990a), Bilevel programming for steady-state chemical process design-I. Fundamentals and Algorithms. *Comp. Chem. Engng.* 14(1): 87.
27. Clark, P.A. and Westerberg, A.W. (1990b), Bilevel programming for steady-state chemical process design-II. Performance study for nondegenerate designs. *Comp. Chem. Engng.* 14(1): 99.

28. Edmunds, T.A. and Bard, J.F. (1991), Algorithms for nonlinear bilevel mathematical programs. *IEEE Trans. on Systems, Man and Cybernetics* 21(1): 83–89.
29. Esposito, W.R. and Floudas, C.A. (1998), Global optimization in parameter estimation of nonlinear algebraic models via the error-in-variables Approach. *Ind. Eng. Chem. Res.* 37: 1841–1858.
30. Falk, J.E. and Liu, J. (1995), On Bilevel Programming. 1. General Nonlinear Cases. *Mathematical Programming* 70: 47–72.
31. Fiacco, A.V. (1976), Sensitivity analysis for nonlinear programming using penalty methods. *Maths. Prog.* 10: 287.
32. Floudas, C.A. (2000), *Deterministic Global Optimization: Theory, Methods and Application*, Vol. 37 of *Nonconvex Optimization and Its Applications*. Kluwer Academic Publishers, Dordrecht, The Netherlands.
33. Floudas, C.A., Pardalos, P.M., Adjiman, C.S., Esposito, W.R., Gümüş, Z.H., Harding, S.T., Klepeis, J.L., Meyer, C.A. and Schweiger, C.A. (1999), *Handbook of Test Problems in Local and Global Optimization*, Vol. 33 of *Nonconvex Optimization and Its Applications*. Kluwer Academic Publishers, Dordrecht, The Netherlands.
34. Floudas, C.A. and Zlobec, S. (1998), Optimality and duality in parametric convex lexicographic programming. In: Migdalas, A., Pardalos, P.M. and Värbrand, P. (eds), *Multilevel Optimization: Algorithms and Applications*, Vol. 20 of *Nonconvex Optimization and Its Applications*. Kluwer Academic Publishers, Dordrecht, The Netherlands. pp. 359–379.
35. Grossmann, I.E. and Floudas, C.A. (1987), Active constraint strategy for flexibility analysis in chemical processes. *Comp. Chem. Engng.* 11(6): 675.
36. Gümüş, Z.H. and Ciric, A.R. (1997), Reactive distillation column design with vapor/liquid/liquid equilibria. *Comp. Chem. Engng.* 21(S): S983.
37. Hansen, P., Jaumard, B. and Savard, G. (1992), New branch and bound rules for linear bilevel programming. *SIAM Journal on Scientific and Statistical Computing* 13: 1194–1217.
38. Hobbs, B.F. and Nelson, S.K. (1992), A nonlinear bilevel model for analysis of electric utility demand-side planning issues. *Annals of Operations Research* 34: 255.
39. Ierapetritou, M.G. and Pistikopoulos, E.N. (1996), Batch plant design and operations under uncertainty. *I&EC Res.* 35: 772.
40. Ishizuka, Y. and Aiyoshi, E. (1992), Double penalty method for bilevel optimization problems. *Annals of Operations Research* 34: 73–88.
41. Jeroslow, R.G. (1985), The polynomial hierarchy and a simple model for competitive analysis. *Math. Programming* 32: 146.
42. Kim, I., Liebman, M.J. and Edgar, T.F. (1990), Robust error-in-variables estimation using nonlinear programming techniques. *AIChE J.* 36(7): 985–993.
43. Kolstad, C.D. (1985), A review of the literature on bilevel mathematical programming. Technical Report LA-10284-MS, US-32, Los Alamos National Lab.
44. LeBlanc, L.J. and Boyce, D.E. (1986), A bilevel programming algorithm for exact solution of the network design problem with user optimal flows. *Trans. Res. B.* 20B(3): 259.
45. Luh, P.B., Ho, Y. and Muralidaran, R. (1982), Load adaptive pricing: an emerging tool for electric utilities. *IEEE Trans. Auto. Contr.* AC-27(2): 320.
46. Luo, Z.Q., Pang, J.S. and Ralph, D. (1997), *Mathematical Programs with Equilibrium Constraints*. Cambridge University Press.
47. Maranas, C.D. and Floudas, C.A. (1994), A Global Optimization Method for Weber's Problem with Attraction and Repulsion. In: Hager, W.W., Hearn, D.W. and Pardalos, P.M. (eds), *Large Scale Optimization: State of the Art*. Kluwer Academic Publishers, Dordrecht, The Netherlands, p. 259.
48. Maranas, C.D. and Floudas, C.A. (1995), Finding all solutions of nonlinearly constrained systems of equations. *Journal of Global Optimization* 7(2): 143–182.

49. Marcotte, P. and Zhu, D.L. (1996), Exact and Inexact Penalty methods for the generalized bilevel programming problem. *Mathematical Programming* 74, 141–157.
50. Migdalas, A. (1995), Bilevel programming in traffic planning: models, methods and challenge. *Journal of Global Optimization* 7: 381.
51. Migdalas, A., Pardalos, P.M. and Värbrand, P. (1998), *Multilevel optimization: Algorithms and Applications*, Vol. 20 of *Nonconvex Optimization and Its Applications*. Kluwer Academic Publishers, Dordrecht, The Netherlands.
52. Onal, H., Darmawan, D.H. and Johnson III, S.H. (1995), A multilevel analysis of agricultural credit dstribution in East Java, Indonesia. *Comp. Op. Res.* 22(2): 227.
53. Pardalos, P.M. and Deng, X. (1997), Complexity Issues in Hierarchical Optimization. *DIMACS Series, Americal Mathematical Society* 37: 219–224.
54. Sahin, K.H. and Ciric, A.R. (1998), A dual temperature simulated annealing approach for solving bilevel programming problems. *Comp. Chem. Engng.* 23: 11–25.
55. Savard, G. and Gauvin, J. (1994), The steepest descent direction for the nonlinear bilevel programming problem. *Operations Research Letters* 15: 265–272.
56. Schweiger, C.S. and Floudas, C.A. (1998), *MINOPT: A modeling language and algorithmic framework for linear, mixed-integer, nonlinear, dynamic and mixed-integer nonlinear optimization*, Vol. Version 3.1. Princeton University.
57. Shimizu, K., Ishizuka, Y. and Bard, J.F. (1997), *Nondifferentiable and Two-Level Mathematical Programming*. Kluwer Academic Publishers, Dordrecht, The Netherlands.
58. Simaan, M. and Cruz, J.B. (1973), On the Stackelberg strategy in nonzero-sum games. *Journal of Optimization Theory and Applications* 11(5): 533.
59. Tuy, H., Migdalas, A. and Värbrand, P. (1993a), A global optimization approach for the linear two-level program. *Journal of Global Optimization* 3: 1.
60. Tuy, H., Migdalas, A. and Värbrand, P. (1993b), A quasiconcave minimization method for solving linear two-level programs. *Journal of Global Optimization* 4: 243.
61. Vicente, L. and Calamai, P. (1994), Bilevel and Multilevel Programming: A bibliography review. *Journal of Global Optimization* 5(2).
62. Vicente, L., Savard, G. and Judice, J. (1994), Descent approaches for quadratic bilevel programming. *Journal of Optimization Theory and Applications* 81(2): 379.
63. Viswanathan, J. and Grossmann, I.E. (1990), *DICOPT++: A Program for Mixed Integer Nonlinear Optimization, User's Guide*. Engineering Design Research Center, Carnegie Mellon University.
64. Visweswaran, V., Floudas, C.A., Ierapetritou, M.G. and Pistikopoulos, E.N. (1996), A decomposition-based global optimization approach for solving bilevel linear and quadratic programs. In: Floudas, C.A. and Pardalos, P.M. (eds), *State of the Art in Global Optimization*, Nonconvex Optimization and its Applications. Kluwer Academic Publishers, Dordrecht, The Netherlands, p. 139.
65. von Stackelberg, H.: (1952), *The Theory of Market Economy*. Oxford University Press.
66. White, D.J. and Anandalingam, G. (1993), A penalty function approach for solving bilevel linear programs. *Journal of Global Optimization* 3: 397–419.
67. Yang, H. and Yagar, S. (1995), Traffic Assignment and Signal Control in Saturated road networks. *Trans. Res. A.* 29A(2): 125.