A Global Optimization Method for Nonlinear Bilevel Programming Problems

Mahyar A. Amouzegar, *Member, IEEE*

***Abstract*—Nonlinear two-level programming deals with optimization problems in which the constraint region is implicitly determined by another optimization problem. Mathematical programs of this type arise in connection with policy problems to which the Stackelberg leader–follower game is applicable. In this paper, the nonlinear bilevel programming problem is restated as a global optimization problem and a new solution method based on this approach is developed. The most important feature of this new method is that it attempts to take full advantage of the structure in the constraints using some recent global optimization techniques.**

***Index Terms*—DC programming, global optimization, nonlinear bilevel optimization, Stackelberg game.**

I. Introduction

The bilevel programming problem (BLPP) is a mathematical model of the leader-follower game. In this game, the control of decision variables are partitioned amongst two players; the leader and the follower. Each player seeks to optimize her objective function. The basic leader/follower strategy was originally proposed for a duopoly by Von Stackelberg [22] in which decisions are made sequentially and cooperation is not allowed. Perfect information is assumed in the sense that both players know the objective functions and the allowable strategies of the other.

The leader moves first by choosing a vector $x \in X \subset R^{n_1}$ in an attempt to optimize her objective function $F(x, y)$. The leader's choice of strategies affects both the follower's objective and decision space. The follower observes the leader's choice and reacts by selecting a vector $y \in Y \subset R^{n_2}$ that optimizes her objective function $f(x, y)$. In doing so, the follower affects the leader's outcome.

It is important to realize the distinction between the bilevel programming problem and the common decomposition of large planning problems into multilevel problems (e.g., [12]). These methods are all concerned with breaking down a large math program into a number of smaller, more tractable units. An unique objective function is used to express the overall system goals. Separate solutions are obtained for each lower-levels and then combined in a master program to yield a complete solution. The basic distinction of this approach from bilevel programming is the assumption that a single objective function can be devised to accurately represent the upper-level as well as the lower-level goals. Even if this objective function can be decomposed, it is highly unlikely that a satisfactory weighing scheme can be developed to make it agreeable to all subdivisions.

Manuscript received January 8, 1998; revised March 26, 1999. This paper was recommended by Associate Editor Y. Y. Haimes.

The author is with RAND Corporation, Santa Monica, CA 90407 USA (e-mail: mahyar@rand.org).

Publisher Item Identifier S 1083-4419(99)09457-1.

The BLPP can be formulated as follows:

$$
\begin{aligned}
\min_{x,y} \quad & F(x,y) \\
\text{s.t.} \quad & x \in X = \{x | G(x) \geq 0\} \\
\text{where } y \text{ solves,} \quad & \qquad\qquad\qquad \text{(P)} \\
& \min_{z} \quad f(x,z) \\
& \text{s.t.} \quad g(x,z) \geq 0 \\
& \qquad\; z \in Y = \{y | H(y) \geq 0\}
\end{aligned}
$$

where G, H, and g are vector valued functions of dimension n_1, n_2, and n_3, respectively. F and f are real-valued functions of appropriate dimensions. The sets of X and Y may represent upper and lower bounds on elements of the vectors x and y.

Bilevel Programming has wide applicability in the area of network design [16], transport system planning [18], management [17], and economics [7], in particular central economic planning. It is, therefore, natural that the theoretical discussion of this paper has been strongly motivated by real-world problems of *Regional Hazardous Waste Management* [3], [5]. In recent years, a great deal of progress has been made in developing algorithms for linear bilevel programming problems such as implicit enumeration [9], K.K.T. approach [11], branch and bound [6], [13], penalty method [4], and many others [21]. The area of nonlinear bilevel programming is still in its infancy with a hand full of algorithms such as branch and bound [8], [10] or global optimization [2], [20].

II. Bilevel Programming Problem

Consider a bilevel hierarchical system where the higher-level decision maker, the leader, controls decision vector $x \in X \subset R^{n_1}$, and the lower-level decision maker, the follower, controls $y \in Y \subset R^{n_2}$. The leader makes her decision first by selecting an $x \in \Omega(X) \subseteq X$, and the follower observing the leader's decision, responds by selecting a decision vector $y \in Y \cap \Omega(x)$. The sets $\Omega(X)$ and $\Omega(x)$ place additional restrictions on the feasible regions of the leader and the followers, respectively, and will be described fully next.

In order to facilitate further discussion of the properties of BLPP, the following definitions are introduced. This notation follows [8].

Let

$$\Omega = \{(x, y) \colon x \in X, y \in Y, g(x, y) \geq 0\}$$

denote the *BLPP's constraint region*. This region includes all the possible combination of choices the leader and the follower may make.

Let

$$\Omega(x) = \{y\colon y \in Y,\ g(x,y) \geq 0\}$$

be *the follower's feasible region for fixed* x. Of course the leader has to choose an x that will not violate the follower's feasible region.

Let

$$\Omega(X) = \{x \in X | \exists y \ni (x,y) \in \Omega\}$$

be *the projection of* Ω *onto the leader's decision space*, and let $M(x)$ be the follower's *rational reaction set* to a given x.

$$M(x) = \{y | y \in \text{ argmin } \{f(x,z) | z \in \Omega(x)\}\}$$

The rational reaction set is an implicit mapping that takes a point, $x \in X$, into a subset of follower's feasible region on which the function f is minimized with respect to $y \in Y$. It should be noted that the follower's problem may be infeasible for certain values of $x \in X$. Therefore, the rational reaction set may be empty for some x.

The leader, by its various choices of x, elicits different rational reaction from the follower. The union of all possible vectors that the leader may select, x, and the corresponding rational reaction set, $y \in M(x)$, is called *the inducible region*. Let IR denote the inducible region defined by

$$IR = \{(x,y)\colon x \in \Omega(X),\ y \in M(x)\}.$$

The leader's problem is then to optimize its objective function over the inducible region

$$\min(F(x,y)\colon (x,y) \in IR).$$

In order to assure that BLPP is well posed we make the assumption that Ω is nonempty and compact. We also assume that the leader will restrict its selection to the set $\Omega(X)$, which guarantees that $M(x) \neq \phi$.

The lower optimization problem, L(x), can be described as follows, let $\sigma(x)$ denote the optimal value of the lower-problem

$$\text{L}(x)\colon \quad \sigma(x) = \min \{f(x,y) | g(x,y) \geq 0, y \in Y\}$$

where f is the objective function of the lower-problem. Further, denote the optimal solution of the lower problem for a given x by $y(x)$.

Proposition 1: Let $\Omega(X)$ be a convex set. If f is a convex function then σ is a convex function on $\Omega(X)$. □

Proposition 2: The bilevel programming problem (P) is equivalent to the following Difference of two Convex functions (DC) problem

$$\begin{aligned} \min \quad & F(x,y) \\ \text{s.t.} \quad & \sigma(x) \geq f(x,y) \\ & (x,y) \in \Omega \end{aligned} \qquad \text{(Q)}$$

where $\Omega = \{(x,y) | x \in X, y \in Y, g(x,y) \geq 0\}$. □

Methods for solving DC problems have been developed in recent years by several authors [15]. However, since problem (Q) has a specific structure, to solve it efficiently, it is important to devise a method that could take advantage of this structure.

In the Section III, we present the basic results needed to develop our method.

III. Preliminaries

The case where the functions, F, f, G, H, and g are linear have been addressed extensively. For details, see [21]. Although the basic proposition that an optimal solution of a linear bilevel program is at a vertex of Ω does not hold in the nonlinear case, there are several properties for the case where the functions are convex. We will explore some of these properties next. Assume F, $-$G, and $-$H are convex in all their arguments, $-$g is convex in y for x fixed, f is strictly convex in y, and that all functions are twice differentiable.

Proposition 3: If all functions are twice differentiable in problem (P) then the inducible region is connected. □

Proof: Since F, $-G$, $-H$, and $-g$ are all convex, and twice differentiable then under the further assumption that f is strictly convex the lower problem will have a unique solution for a given $x \in \Omega(X)$. Then $M(x)$ is single valued and the inducible region could be replaced by a unique response function. Therefore $M(x)$ is a point-to-point mapping and continuous. By [14, Corollary 8.1], IR is connected. □

Corollary 1: If f is quadratic and Ω is polyhedral then the inducible region is piece-wise linear.

Proof: See [8, Corollary 1]. □

Proposition 4: Let $f(x,y)$ be quadratic in (x,y) and Ω be a nonempty convex polytope, then the convex hull of the set of solutions to problem (P) is a convex polytope.

Proof: Denote $L = \{(x,y) \in R^{n_1+n_2} | \sigma(x) - f(x,y) \geq 0\} \neq \phi$. Let $(\overline{x},\overline{y}) \in IR$ and $(\overline{x},\overline{y}) \notin int(conv(L))$. Then by the separating hyperplane theorem there is a nonzero vector p such that $p((x,y) - (\overline{x},\overline{y})) \leq 0$ for each $(x,y) \in L$. Assume $H(\overline{x},\overline{y}) = \{(x,y) \in \Omega | p((x,y) - (\overline{x},\overline{y})) \leq 0\}$ and $(\overline{x},\overline{y}) \in H(\overline{x},\overline{y})$. Therefore $\Omega \cap H(\overline{x},\overline{y})$ is a convex polytope. Let $E_i, i = 1,\cdots,K$ denote the edges of Ω that contains the vertices of $\Omega \cap H(\overline{x},\overline{y})$. $(\overline{x},\overline{y})$ is a convex combination of these vertices, which must lie amongst $E_i \cap H(\overline{x},\overline{y}) \subset E_i \cap L$, for $i = 1,\cdots,K$.

Now

$$(\overline{x},\overline{y}) \in conv \left\{ \bigcup_i \{E_i \cap L\} \right\}.$$

Therefore

$$\begin{aligned} conv \{\Omega \cap L\} &= conv \left\{ \bigcup_i \{E_i \cap L\} \right\} \\ &= conv \left\{ \bigcup_i conv \{E_i \cap L\} \right\}. \end{aligned}$$

Each $conv\{E \cap L\}$ is a convex polytope and the convex hull of a finite union of convex polytope is a convex polytope. □

Proposition 5: If there is a $\hat{x} \in int(\Omega(X))$ so that $\sigma(\hat{x}) > -\infty$ then $\sigma(x) > -\infty$ for all $x \in \Omega(X)$ □

IV. Procedure

We present a procedure based on branch and bound method where the branching is performed on the leader's decision variables with the use of simplicial subdivision. The bounding is based on outer approximation scheme on the convex constraint.

Let

$$\sigma(v^i) = \alpha v^i + \beta, \quad \alpha \in R^{n_1}, \quad \beta \in R^1, \qquad i = 0, \cdots, n_1$$

with α and β uniquely determined. Denotes the n_1-simplex with vertices, $v^\circ, \cdots, v^{n_1}$ by $\Delta = conv\{v^\circ, \cdots, v^{n_1}\}$. Hence, $\psi(x) = \alpha x + \beta$ with $x \in \Delta$ is the unique affine function going through $n+1$ linearly independent vectors. Consequently

$$x = \sum_{i=0}^{n_1} \lambda_i v^i, \quad \sum_{i=0}^{n_1} \lambda_i = 1, \qquad \lambda_i \geq 0$$

and from the convexity of σ it follows that

$$\psi(x) = \sum_{i=0}^{n_1} \lambda_i \psi(v^i) = \sum_{i=0}^{n_1} \lambda_i \sigma(v^i) \geq \sigma(x).$$

Hence, $\psi(x) \geq \sigma(x)$ for all $x \in \Delta$, and $\psi(x)$ is an unique convex subfunctional of $\sigma(x)$ on Δ.

We propose a branch and bound method on the simplex Δ where the branching is performed with respect to x-variable by subdividing the space into subsets of the form $\Delta \times Y$ with Δ a n_1-simplex. The bounding is based on solving a relaxed problem at each iteration on the simplex Δ. The relaxed problem is obtained by the use of the convex subfunctional. At iteration k, we have a number of subsimplices, $\mathcal{S}$, that remain to be examined. For each simplex in this collection we shall solve the following convex program:

$$\begin{aligned} \min \quad & F(x,y) \\ \text{s.t.} \quad & (x,y) \in \Omega \\ & \psi(x) \geq f(x,y) \qquad \text{CP}(\Delta) \\ & x \in \Delta. \end{aligned}$$

Let (x^k, y^k) denote the solution to CP(Δ). If $F(x^k, y^k) \geq F(x^\star, y^\star)$ (where $(x^\star, y^\star)$ is the current best solution), then the current simplex Δ is discarded from further consideration. The new simplex Δ_k is chosen for branching according to the minimal value of $F(x,y)$. To subdivide Δ_k we can follow several alternative rules, but in order to guarantee convergence we have chosen the following rule based on [19]. Let $\Delta_k = [v^{k_\circ}, \cdots, v^{k_{n_1}}]$ and $\mathcal{I}(\Delta)$ denoting the generating index of Δ with $\mathcal{I}(\Delta_\circ) = 0$, $\mathcal{I}(\Delta') = \mathcal{I}(\Delta) + 1$ whenever Δ' is a child of Δ. If $\max(\|x^k - v^{k_r}\|_{r=0,\cdots,n_1}) \leq \varepsilon d(\Delta_k)$ (where $d(\Delta)$ denotes the diameter of Δ, i.e., the length of its longest edge) and $\mathcal{I}(\Delta_k)$ is not a multiple of N, then divide Δ_k with respect to x^k; otherwise, divide with respect to the midpoint of a longest edge. ε and N are pre-chosen parameter with N a natural number experimentally set at five or less $(N \leq 5)$ and $\varepsilon \in (0,1)$.

The current best solution is updated according to the bilevel feasible solution. If at iteration k, $(x^k, y^k) \in IR$ and $F(x^k, y^k) < F(x^\star, y^\star)$ then update $(x^\star, y^\star)$. More often than not, the solution (x^k, y^k) will not be bilevel feasible and the updating is performed by solving the lower problem at x^k.

We end this section with a simple illustrative example borrowed from [8].

A. Example 1

$$\begin{aligned} \min \quad & (x-5)^2 + (2y+1)^2 && (1) \\ \min \quad & (y-1)^2 - 1.5xy && (2) \\ & 3x - y \geq 3 && (3) \\ & -x + .5y \geq -4 && (4) \\ & -x - y \geq -7. && (5) \end{aligned}$$

We solve the relaxed problem by dropping the follower's objective function (2) to obtain an initial solution$(x^\circ, y^\circ) = (4,0)$ with an upper objective value of 2. Fixing $x = 4$ and solving the lower problem (i.e., (2)–(5)) yields a local solution $(x^\star, y^\star) = (4,3)$ with the current best objective value at $F^\star = 50$. Compute $\psi(x) = -3.75x + 4.75$ with $\Delta_\circ = [1,5]$. Solving CP($\Delta_\circ$) produces a bilevel infeasible solution $(x^1, y^1) = (1.440\,25, .444\,264\,6)$ but solving the lower problem at x^1 produces a better feasible solution. Updating $(x^\star, y^\star) = (1.440\,25, 1.320\,75)$ with $F^\star = 25.932\,34$. Subdividing $\Delta_\circ$ with respect to the x variable generates two simplices $\Delta_1 = [1, 1.440\,250]$ and $\Delta_2 = [1.440\,250, 5]$.

Solving CP(Δ_1) with $\psi(x) = -8.518\,876x + 9.518\,876$ produces a bilevel feasible solution $(x^2, y^2) = (1,0)$ with an objective value of $F^\star = 17$. Update $(x^\star, y^\star) = (x^2, y^2) = (1,0)$.

Solving CP(Δ_2) with $\psi(x) = -3.160\,212x + 1.801\,06$ generates an inferior solution $(x^3, y^3) = (4.646\,558\,1.988\,632)$ with objective value of 24.898 08. Δ_2 is then fathomed and the algorithm terminates with a global solution of $(x^\star, y^\star) = (1,0)$.

B. Further Refinement and Extensions

We may further refine this method by enclosing the convex constraint by a polyhedron. This is achieved by replacing the convex constraint by a series of affine functions $\ell_j(x,y) \geq 0$, $j \in J_k \subset \{1, 2, \cdots, k\}$ that contains Ω. If a given solution $(x^k, y^k) \in \Omega$ then we set $J_{k+1} = J_k$ (i.e., no change). Otherwise, we set $J_{k+1} = J_k \cup \{k+1\}$ and construct a new affine function $\ell_{k+1}(x,y) \geq 0$ strictly separating (x^k, y^k) from Ω, i.e., $\ell_{k+1}(x^k, y^k) < 0$, $\ell_{k+1}(x,y) \geq 0$, for all $(x,y) \in \Omega$. Such hyperplanes are easily constructed by

$$\ell_{k+1}(x,y) = p^k \cdot (x - x^k) + q^k \cdot (y - y^k) + g(x^k, y^k)$$

where $p^k = \nabla_{x^k} g(x,y)$ and $q^k = \nabla_{y^k} g(x,y)$. Note that for the sake of notational simplicity we have assumed that $\Omega = \{(x,y) | g(x,y) \geq 0\}$ (i.e., all the convex constraints are functions of both variables).

Now, at iteration k, we have a linear system comprised of the collection of ℓ_j, $j \in I_k$ with $I_k \subset \{1, \cdots, k\}$ such that it contains Ω and a number of subsimplices, $\mathcal{D}$ that remains to be examined. The relaxed problem is therefore

$$\begin{aligned} \min \quad & F(x,y) \\ \text{s.t.} \quad & \ell_i(x,y) \geq 0, \qquad i \in I_k \\ & \psi(x) \geq f(x,y) \qquad \text{RP}(\Delta) \\ & x \in \Delta. \end{aligned}$$

As before the branching is performed according to the minimal value of $F(x,y)$.

Algorithm:

Initialization:

Let (x°, y°) be an optimal solution of (C), the convex problem
obtained from (Q) by dropping the DC constraint,

$$\begin{aligned} \min \quad & F(x,y) \\ & (x,y)\in\Omega \qquad \text{(C)} \end{aligned}$$

If $\|\sigma(x^\circ) - f(x^\circ, y^\circ)\| \le \epsilon$ for ϵ small then stop and (x°, y°) is a
trivial solution to (Q) and therefore to (P).
Otherwise
Let $F^\star = F(x^\circ, y(x^\circ))$ if $y(x^\circ)$ exists, otherwise let
$F^\star = +\infty$ denote the current best value.
Set $I = \phi, \varrho_1 = \mathcal{S}_1 = \{\Delta_\circ\}$, where $\Delta_\circ$ is an n_1-simplex containing the projection of Ω on R^{n_1}.
Set $k = 0, \wp_k = \{\Delta_\circ\}$.

While $\wp_k \neq \phi$ **do**
For each $\Delta \in \mathcal{S}_k$, let $(x_\Delta, y_\Delta) \in \Omega$ (if exists) be a solution of RP(Δ)
with its optimal value $F^\star_\Delta$.
If for any $\Delta \in \mathcal{S}_k, F^\star_\Delta > F^\star$ then delete Δ.
Let $(x_\Delta, y(x_\Delta)) \in IR$ be the bilevel feasible solution
of L(x_Δ).
Update $F^\star$ and $(x^{\star^k}, y^{\star^k})$
Let $\wp_k$ be the collection of remaining simplices.
Select $\Delta_k \in argmin\{F^\star_\Delta | \Delta \in \wp_k\}$, if $F^\star_\Delta$ exists.
Otherwise select the simplex with the least infeasible solution.
Subdivide Δ_k and denote ϱ_{k+1} the partition of Δ_k.
Let (x^k, y^k) be an optimal solution to RP(Δ_k).
If $(x^k, y^k) \in \Omega$ then set $I_{k+1} = I_k$
else
Set $I_{k+1} = I_k \cup \{k+1\}$.
Construct $\ell_{k+1}(x,y)$.
Set $\mathcal{S}_{k+1} = (\wp_k / \{\mathcal{S}_k\}) \cup \varrho_{k+1}$.
$k \leftarrow k+1$.
End While
Output $F^\star$ and $(x^{\star^k}, y^{\star^k})$.

Lemma 1: Every accumulation point of the sequence $\{(x^k, y^k)\}$ is feasible for (P).

Proof: See [15, Theorem II.2.]. □

Theorem 1: If the algorithm is infinite, then $(x^\star, y^\star) = \lim_{k\to\infty} (x^{\star^k}, y^{\star^k})$ solves (P).

Proof: By Lemma 1, $(x^\star, y^\star)$ is feasible. If (x,y) is feasible and $x \in \Delta_i$, where Δ_i is deleted for $i < j$ then we have $F(x,y) \ge F^\star_\Delta \ge F^\star_{\Delta_i} \ge F^\star_{\Delta_j}$. On the other hand if $(x,y) \in \wp_k$ then $F(x,y) \ge F^\star_\Delta \ge F^\star_{\Delta_j}$. In either case, $F(x,y) \ge F^\star_{\Delta_j}$ for all k and therefore, $F(x,y) \ge F\star$. □

An obvious extension to this problem is when $F(x)$ is concave or DC in the sense that the objective of the upper level is of the form, $\mathcal{F}(x,y) - F(x)$, where both $\mathcal{F}$ and F are convex. Clearly the problem

$$\begin{aligned} \min \quad & \mathcal{F}(x,y) - F(x) \\ & (x,y)\in\Omega \end{aligned}$$

can be converted to concave minimization problem by rewriting the problem as

$$\min\{t - F(x) | (x,y) \in \Omega, \mathcal{F}(x,y) \le t, \gamma \le t \le \mu, \}$$

where $\gamma = \min\{\mathcal{F}(x,y) | (x,y) \in \Omega\}$, and $\mu = \max\{\mathcal{F}(x,y) | (x,y) \in \Omega\}$. Therefore the problem we are interested is of the form

$$\begin{aligned} \min \quad & F(x) \\ \text{s.t.} \quad & (x,y)\in\Omega \qquad \text{(QC)} \\ & \sigma(x) \ge f(x,y) \end{aligned}$$

where as before $f(x,y)$ denotes the objective function of the lower problem, which is either affine or convex.

Definition 1: $\varphi(x)$ is defined to be the lower subfunctional of $F(x)$ on a n_1-simplex Δ, if $\varphi(x)$ is affine and

$$\varphi(v^i) = f(v^i), \qquad i = 0, 1, \cdots, n_1$$

where $v^\circ, v^1, \cdots, v^{n_1}$ are the vertices of the n_1-simplex Δ. The problem RP(Δ) can be reformulated as follows:

$$\begin{aligned} \min \quad & \varphi_\Delta(x) \\ \text{s.t.} \quad & \ell_i(x,y) \ge 0, \qquad i \in I_k \\ & \psi(x) \ge f(x,y) \\ & x \in \Delta. \end{aligned}$$

With minor modification to the algorithm, we can solve (QC) with similar results.

We end this section with another example from [8].

C. Example 2

$$\begin{aligned} \min \quad & -x_1^2 - 3x_2 - 4y_1 + y_2^2 \\ \text{s.t.} \quad & x_1^2 + 2x_2 \le 4 \\ & x_1, x_2 \ge 0 \\ & \min\ 2x_1^2 + y_1^2 - 5y_2 \\ & \text{s.t.}\ x_1^2 - 2x_1 + x_2^2 - 2y_1 + y_2 \ge -3 \\ & \qquad x_2 + 3y_1 - 4y_2 \ge 4 \\ & \qquad y_1, y_2 \ge 0. \end{aligned}$$

To begin, the relaxed problem yields $x^\circ = (0,2)$, $y^\circ = (4,1)$ with an objective value of -21. Solving the lower bound at $x^\circ = (0,2)$ gives a bilevel feasible solution of $(x^\circ, y(x^\circ)) = ((0,2), (1.875, .9063))$ with an upper bound $F^\star = -12.68$

(which, incidentally, has been incorrectly reported in [8] as -14.13). Let the initial simplex be

$$\Delta_\circ = \{v^\circ = (0,0)^\top, v^1 = (2.5,0)^\top, v^2 = (0,2.5)^\top\}$$

with $\psi(x) = 4.970x_1 - 1.280x_2 + 1.560$. Solving

$$\begin{aligned} \min \quad & -x_1^2 - 3x_2 - 4y_1 + y_2^2 \\ & 4.970x_1 - 1.280x_2 + 1.560 \geq 2x_1^2 + y_1^2 - 5y_2 \\ & \quad \mathrm{RP}(\Delta_\circ) \\ & x \in \Delta_\circ \end{aligned}$$

gives the following infeasible solution: $(x_1, x_2) = (1.844, .6556)$ and $(y_1, y_2) = (3.274, 1.527)$. Accordingly we divide the initial simplex into three subsimplices at with respect to $v^3 = (1.844, .6556)$: $\Delta_{11} = \{v^\circ, v^2, v^3\}$, $\Delta_{12} = \{v^\circ, v^1, v^3\}$, $\Delta_{13} = \{v^1, v^2, v^3\}$ and compute $\psi(x)$ for each subsimplices. Let $\mathcal{S}_1 = \{\Delta_{11}, \Delta_{12}, \Delta_{13}\}$ and $\ell_1 = -3.688x_1 - 2x_2 + 7.402$.

No feasible solution is obtained by solving RP(Δ) for each $\Delta \in \mathcal{S}_1$.

Δ_{11} is further subdivided with respect to the solution of RP(Δ_{11}), $v^4 = x = (.4839, 2.016)$ and $y = (2.956, 1.691)$: $\Delta_{21} = \{v^\circ, v^2, v^4\}$, $\Delta_{22} = \{v^\circ, v^3, v^4\}$ and $\Delta_{23} = \{v^2, v^3, v^4\}$ with $\ell_2 = -.9678x_1 - 2x_2 + 4.234$.

Let $\mathcal{S}_2 = \{\Delta_{12}, \Delta_{13}, \Delta_{21}, \Delta_{22}, \Delta_{23}\}$ and solve RP(Δ) for each $\Delta \in \mathcal{S}_2$. None of the subsimplices produce a feasible solution and therefore further subdivision is required. In fact, no feasible solution is obtained until an additional four subdivisions are performed. The following hyperplanes where also introduced at steps 3–7, respectively. $\ell_3 = -2x_2 + 4$, $\ell_4 = x_2 + 3y_1 - 4y_2 - 4$, $\ell_5 = -.4839x_1 - 2x_2 + 4.059$, $\ell_6 = -.1579x_1 - 2x_2 + 4.006$ and $\ell_7 = -.154x_1 - 2x_2 + 4.006$.

At iteration 7, we have the following subsimplices at hand:

$$\mathcal{S}_7 = \{\Delta_{12}, \Delta_{13}, \Delta_{22}, \Delta_{23}, \Delta_{31}, \Delta_{41}, \Delta_{52}, \Delta_{71}, \Delta_{72}, \Delta_{73}\}$$

with $\Delta_{31} = \{v^\circ, v^4, v^5 = (0, 1.25)^\top\}$, $\Delta_{41} = \{v^4, v^5, v^6 = (0, 1.875)^\top\}$, $\Delta_{52} = \{v^4, v^6, v^7 = (.2419, 2.258)^\top\}$, $\Delta_{71} = \{v^6, v^8 = (0, 2.188)^\top, v^9 = (.0771, 1.997)^\top\}$, $\Delta_{72} = \{v^7, v^8, v^9\}$ and $\Delta_{73} = \{v^6, v^7, v^9\}$. The simplex $\Delta_{62} = \{v^2, v^7, v^8\}$ was fathomed since RP(Δ_{62}) was infeasible. Solving RP(Δ) for each $\Delta \in \mathcal{S}_7$ will allow us to delete Δ_{72} and Δ_{73} due to infeasibility and yields the first feasible solution in the subsimplex Δ_{71}.

The simplex Δ_{71} is further subdivided into: $\Delta_{81} = \{v^6, v^8, v^{10} = (.0370, 1.997)^\top\}$, $\Delta_{82} = \{v^6, v^9, v^{10}\}$ and $\Delta_{83} = \{v^8, v^9, v^{10}\}$. Solving the RP($\Delta$) over all the remaining simplices gives feasible solutions for the last three subsimplices, Δ_{81}, Δ_{82} and Δ_{83} with optimal values of $F^\star_{\Delta_{81}} = -12.69$, $F^\star_{\Delta_{82}} = -12.75$ and $F^\star_{\Delta_{83}} = -12.67$, respectively. Since $F^\star_{\Delta_{83}} > F^\star$, the simplex Δ_{83} is fathomed. Next, Δ_{82} is selected for subdivision that produces a new hyperplane $\ell_8 = x_2 + 3y_1 - 4y_2 - 4$, which consequently leads to an optimal value greater than the current best value and is deleted from further consideration.

The simplex Δ_{81} is selected for subdivision with its children: $\Delta_{91} = \{v^8, v^{10}, v^{11} = (0, 2.031)^\top\}$ and $\Delta_{92} = \{v^6, v^{10}, v^{11}\}$. $F^\star_{\Delta_{91}} > F^\star$ and is deleted. Solving RP(Δ_{92}) gives the optimal solution: $x = (0, 2)$ and $y = (1.875, .9063)$ that is bilevel feasible with an optimal value of $F^\star = -12.68$.

At iteration 11, we have the following remaining simplices at hand:

$$\mathcal{S}_7 = \{\Delta_{12}, \Delta_{13}, \Delta_{22}, \Delta_{23}, \Delta_{31}, \Delta_{41}, \Delta_{52}\}.$$

The simplex Δ_{12} is subdivided into $\Delta_{12,1} = \{v^\circ, v^3, v^{12} = (2.007, 0)^\top\}$ and $\Delta_{12,2} = \{v^1, v^3, v^{12}\}$. $\Delta_{12,2}$ deleted due to infeasibility. The former adds a new hyperplane $\ell_9 = 1.302x_1 + 1.174x_2 - 2y_1 + y_2 - .070\,46$ to problem RP($\Delta_{12,1}$) that yields a worst bilevel feasible solution than the current best solution.

Solving the problem RP(Δ) for each $\Delta \in \mathcal{S}_7$ gives either infeasible solutions or bilevel feasible solutions that are inferior to the current best solution. Consequently at iteration 12 all other remaining simplices are deleted and the current best solution is declared as an optimal solution.

V. Hazardous Waste Management

We stated, in the introduction section, that the theoretical discussion of this paper has been motivated by a real-world problem of regional hazardous waste management. In this section, we present a brief description of this problem.[1]

The California legislation allows counties to participate in regional associations for hazardous waste management planning. The two principle associations are the Association of Bay Area Governments (ABAG), comprised of Alameda, Contra Costa, Marin, Napa, San Francisco, San Mateo, Santa Clara, Solano, and Sonoma counties, and the Southern California Hazardous Waste Management Authority (SCHWMA), comprised of Imperial, Los Angeles, Orange, Riverside, San Bernardino, San Diego, and Santa Barbara counties. ABAG and SCHWMA account for approximately 25% and 50%, respectively, of all hazardous waste generation in the state.

In the Northern California model, the leader's role is taken by ABAG and the follower's role by the collection of firms operating in the Bay Area. The leader, in order to encourage source reduction (i.e., reducing waste at the source), may adopt a policy of rewarding firms for each unit of source reduction beyond its specified lower limit. At the same time, it desires to regulate firms who fail to meet the minimum source reduction standard and for shipping hazardous waste to incinerators. The economic approach to hazardous waste control is based on the regulation of the behavior of the firms. Hence, a tax system induces the polluters to reduce their discharge to a level where their marginal cost of a proper treatment (i.e., recycling or source reduction) equals the marginal cost of pollution damage (taxes can be used as a surrogate measure of pollution damage). Beyond this level, it is cheaper to pay the tax than continue the treatment process and at optimal tax rate, the cost of any pollution related damage is totally internalized.

The firms, of course, incur other costs other than the penalty (tax) set by the government. The firms, in planning their waste management policy, need to consider such costs as the on-site recycling cost (including the setup and operation costs), off-site recycling costs, and incineration costs. This type of

[1] For a more detailed description of hazardous waste management, see [5].

interaction leads to an hierarchical decision making where the government (e.g., ABAG) assumes the role of the leader that makes decision on prices and taxes. On the other hand, the firms observing the decision made by the Central Authority (CA) set their own allocation polices.

A. Conceptual Model

We modeled the above problem as a bilevel programming problem. The leader's goal was to minimize the total cost of transportation, setup costs and the cost of firms using the off-site incinerators, recycling or disposal sites. Due to the economies-of-scale many of the cost functions were nonlinear and concave. The decision variables for the leader consisted of the unit price for the off-site incinerator or recycling facilities, taxes levied based on waste type and other transportation decisions. Prices and taxes are the charges that the follower has to pay per unit of waste treatment and/or discharge.

The firms faced with the taxes and charges want to minimize their total operating cost. Their decision variables consist of the amount of waste sent for recycling, incineration or disposal.

The leader's constraint is the follower's optimization problem (i.e., $X = \phi$). The follower's constraint includes, on- and off-site capacities, and conservation equations.

The data set that we worked on, contained 20 waste streams, three sizes of incinerators and recycling facilities were made available with about 27 waste generators. Under such limited scenario, the leader had to deal with about 500 continuous and 54 discrete decision variables and the follower's problem had about 5000 continuous and 81 discrete decision variables. Although the number of decision variables seems small for a traditional single level optimization problem, it is immense for this type of nonconvex optimization problem. Neither our algorithm nor any other reported method is able to solve such a problem at this point.

The detailed description of the problem and the solution to its linearized version is reported in [5]. We also solved a severely restricted nonlinear model by reducing the number of decision variables to about 15 for the leader and 15 for the follower. The result is, of course, academic and it is not presented in this paper.

VI. Concluding Remarks

In this paper, we presented a global optimization method for solving nonlinear bilevel programming problems. The bilevel programming problem is converted to an equivalent DC problem with special structure. We proposed a branch and bound method on the simplex that enclosed the convex constraint and developed ideas to extend the method in order to include a more general structure for the original objective functions.

The handful of methods that are available to solve nonlinear bilevel programming problems are either designed to solve more relaxed types of objectives or efficiency was not the goal of their methods. Further, we must stress that solving large scale problems is extremely difficult with the current state of art and the focus has been in developing methods that can in fact guarantee global solution. Keeping that in mind, we compared our method with the barrier method of [1] and the branch and bound method of [8] using the problem sizes reported in the latter paper. We used convex objectives and constraints with up to 30 decision variables. Computational times were not available so we had to rely on the number of iterations, which may not be the best method of gauging computational efficiency. Our method can be considered more "efficient" than the barrier method but at times was not able to perform better than the branch and bound method.

Computational experiments to better test the efficiency of this method with a relatively larger size problem is underway and may help in our development of better subdivisional schemes.

References

[1] E. Aiyoshi and K. Shimizu, "Hierarchical decentralized system and its new solution by a barrier method," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-11, pp. 444–449, 1981.

[2] F. Al-Khayyal, R. Horst, and P. Pardalos, "Global optimization of concave functions subject to quadratic constraints: An application in nonlinear bilevel programming," *Ann. Oper. Res.*, vol. 34, pp. 125–147, 1992.

[3] M. A. Amouzegar and S. E. Jacobsen, "A decision support system for regional hazardous waste management alternatives," *J. Appl. Math. Decis. Sci.*, vol. 2, no. 1, 1998.

[4] M. A. Amouzegar and K. Moshirvaziri, "A penalty method for linear bilevel programming problems," in *Multilevel Optimization: Algorithms, Complexity and Applications*, A. Migdalas, M. Pardalos, and P. Värbrand, Eds. Norwell, MA: Kluwer, 1997, ch. 11.

[5] ——, "Determining optimal pollution control policies: An application of bilevel programming," *Eur. J. Oper. Res.*, vol. 119, no. 1, 1999.

[6] J. Bard and J. Moore, "A branch and bound algorithm for the bilevel programming problem," *SIAM J. Sci. Statist. Comput.*, vol. 11, pp. 281–292, 1990.

[7] J. F. Bard, "Regulating nonnuclear industrial waste by hazard classification," *J. Environ. Syst.*, vol. 13, pp. 21–41, 1983/1984.

[8] ——, "Convex two level programming," *Math. Progr.*, vol. 40, pp. 15–27, 1988.

[9] W. Candler and R. Townsley, "A linear two-level programming problem," *Comput. Oper. Res.*, vol. 9, no. 1, pp. 59–76, 1982.

[10] T. Edmunds and J. Bard, "Algorithms for nonlinear bilevel mathematical programming," *IEEE Trans. Syst., Man, Cybern.*, vol. 21, pp. 83–89, 1991.

[11] J. Fortuny-Amat and B. McCarl, "A representation and economic interpretation of a two-level programming problem," *J. Oper. Res. Soc.*, vol. 32, pp. 783–792, 1981.

[12] A. M. Geoffrion, "Primal resource-directive approaches for optimizing nonlinear decomposable systems," *Oper. Res.*, vol. 18, no. 13, pp. 375–403, 1970.

[13] P. Hansen, B. Jaumard, and G. Savard, "New branch and bound rules for linear bilevel programming," *SIAM J. Sci. Statist. Comput.*, vol. 13, no. 5, pp. 1194–1217, 1992.

[14] W. W. Hogan, "Point-to-set maps in mathematical programming," *SIAM Rev.*, vol. 25, no. 3, pp. 591–603, 1973.

[15] R. Horst and H. Tuy, *Global Optimization*. Berlin, Germany: Springer-Verlag, 1993.

[16] P. Marcotte, "Network optimization with continuous control parameters," *Transp. Sci.*, vol. 17, pp. 181–197, 1983.

[17] T. Miller, T. Friesz, and R. Tobin, "Heuristic algorithms for delivered price spatially competitive network facility location problems," *Ann. Oper. Res.*, vol. 34, pp. 177–202, 1992.

[18] S. Suh and T. Kim, "Solving nonlinear bilevel programming models of the equilibrium network design problem: A comparative review," *Ann. Oper. Res.*, vol. 34, pp. 203–218, 1992.

[19] H. Tuy, "Effect of subdivision strategy on convergence and efficiency of some global optimization algorithms," Tech. Rep., Inst. Math., Hanoi, Vietnam, 1990.

[20] L. Vicente, G. Savard, and J. Júdice, "Descent approach for quadratic bilevel programming," *J. Opt. Theory Applicat.*, vol. 81, pp. 379–399, 1994.

[21] L. N. Vicente and P. H. Calamai, "Bilevel and multilevel programming: A bibliography review," *J. Global Optim.*, vol. 5, no. 3, 1994.

[22] H. von Stackelberg, *The Theory of the Market Economy*. London, U.K.: William Hodge, 1952.

Mahyar A. Amouzegar (M'94) received the Ph.D. degree in operations research from the Electrical Engineering Department, University of California, Los Angeles.

He is an Associate Professor of Computer Information Systems, California State University, Dominguez Hills, and an Operations Research Analyst, RAND, Santa Monica, CA. Previously, he was with the Faculty of Information and Mathematical Sciences, Massey University, New Zealand. His current research includes military policy analysis, logistics, and optimization. He has performed applied research in environmental economics and has done theoretical work in nonconvex optimization. He is the founder and Managing Editor of the *Journal of Applied Mathematics and Decision Sciences.*