JOURNAL OF OPTIMIZATION THEORY AND APPLICATIONS: Vol. 81, No. 2, MAY 1994

Descent Approaches for Quadratic Bilevel Programming[1]

L. VICENTE,[2] G. SAVARD,[3] AND J. JÚDICE[4]

Communicated by R. Sargent

Abstract. The bilevel programming problem involves two optimization problems where the data of the first one is implicitly determined by the solution of the second. In this paper, we introduce two descent methods for a special instance of bilevel programs where the inner problem is strictly convex quadratic. The first algorithm is based on pivot steps and may not guarantee local optimality. A modified steepest descent algorithm is presented to overcome this drawback. New rules for computing exact stepsizes are introduced and a hybrid approach that combines both strategies is discussed. It is proved that checking local optimality in bilevel programming is a NP-hard problem.

Key Words. Bilevel programming, nonconvex and nondifferentiable optimization, quadratic programming, computational complexity.

1. Introduction

The bilevel programming problem can be defined as

$$\min_{x,y} \quad F(x, y),$$
$$\text{s.t.} \quad g(x, y) \leq 0,$$

where y is the solution of the lower level problem

$$\min_{y} \quad f(x, y),$$
$$\text{s.t.} \quad h(x, y) \leq 0,$$

[1]Support of this work has been provided by INIC (Portugal) under Contract 89/EXA/5, by FCAR (Québec), and by NSERC and DND-ARP (Canada).
[2]Lecturer, Departamento de Matemática, Universidade de Coimbra, Coimbra, Portugal.
[3]Professor, Collège Millitaire Royal de St. Jean, St. Jean, Québec, Canada.
[4]Professor, Departamento de Matemática, Universidade de Coimbra, Coimbra, Portugal.

0022-3239/94/0500-0379$07.00/0 © 1994 Plenum Publishing Corporation

and $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$. Here, $x(y)$ is called the upper (lower) variable. In the same way, $F(f)$ is called the upper (lower) level objective function.

Bilevel programming is an important area in nondifferentiable and nonconvex optimization. Several applications of this problem have appeared in the literature (Ref. 1). A number of algorithms have been designed for finding global minima of bilevel programs when all the functions are linear. Among other interesting approaches, the enumerative algorithms of Bard and Moore (Ref. 2) and Hansen, Jaumard, and Savard (Ref. 3) and a sequential linear complementarity problem (LCP) method (Ref. 4) should be distinguished. These procedures can be extended for the linear-quadratic case, where the functions F, g, h are linear but the lower level function f is strictly convex quadratic.

Bilevel programs with strictly convex quadratic lower level problems and nonlinear upper level functions are much more difficult problems. To date, even for specific nonlinear instances of F, only enumerative procedures have been proposed to find a global minimum (Refs. 5–8). Penalty function approaches (Refs. 9–12) and other descent algorithms (Refs. 13 and 14) have also been developed but only guarantee local minima.

Some attention has also been focused on defining optimality conditions for bilevel programming problems. A first attempt is reported in Ref. 15, which exploits an interesting bilevel reformulation involving an infinite and parametric set of constraints. However, this reformulation has a difficult structure, and a counterexample pointed out in Ref. 16 shows the incorrectness of such conditions. Different necessary optimality conditions can be found in Refs. 17–19.

In this paper, we propose two different descent algorithms for the solution of bilevel programs in which the lower level function is strictly convex quadratic, the upper level function is quadratic (strictly convex or concave), and the constraints of the lower level problem constitute a convex polyhedron. A first algorithm is based on modified pivot steps that enforce direct movements along the induced region and may not achieve a local solution in general. This is not the case of concave upper level functions for which it is proved that a local minimum is always reached. A modification of the steepest descent algorithm (Ref. 19) is introduced using the sequential LCP algorithm (Ref. 20) for an efficient computation of each steepest descent direction. The use of the sequential LCP method seems appropriate although it may face difficulties when a local solution is at hand. In fact, we prove that checking (strict or not) local optimality for bilevel programming is a NP-hard problem. An appropriate technique for computing exact stepsizes is presented. A hybrid approach is also proposed which combines both methods and takes advantage of the particular benefits of both strategies.

The structure of the paper is as follows. In Section 2, the quadratic bilevel program is introduced as well as specific definitions and properties. The algorithms are described in Sections 3 and 4, where the particularly strictly convex and concave instances of the upper level function are also discussed. Complexity issues are addressed in Section 5 and some concluding remarks are reported in the last section of this paper.

2. Problem Definition and Properties

The quadratic bilevel programming problem can be stated as follows:

$$\text{(QBP)} \quad \min_{x,y} \quad F(x,y) = 1/2\begin{bmatrix} x \\ y \end{bmatrix}^T \begin{bmatrix} C_1 & C_3 \\ C_3^T & C_2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}^T \begin{bmatrix} x \\ y \end{bmatrix},$$

$$\text{s.t.} \quad x \geq 0,$$

$$y \in \operatorname{argmin}\{f(x,y) = 1/2 y^T Q y + y^T S x + d^T y: \\ Ax + By \leq b, y \geq 0\},$$

where

$$c_1 \in \mathbb{R}^n, \quad c_2, d \in \mathbb{R}^m, \quad C_1 \in \mathbb{R}^{n \times n}, \quad Q, C_2 \in \mathbb{R}^{m \times m},$$

$$S, C_3^T \in \mathbb{R}^{m \times n}, \quad A \in \mathbb{R}^{l \times n}, \quad B \in \mathbb{R}^{l \times m}, \quad b \in \mathbb{R}^l.$$

We assume that

$$C = \begin{bmatrix} C_1 & C_3 \\ C_3^T & C_2 \end{bmatrix}$$

and Q are both symmetric positive-definite matrices. This implies that both the relaxed problem in the variables x and y,

$$\text{(RP)} \quad \min_{x,y} \quad 1/2\begin{bmatrix} x \\ y \end{bmatrix}^T \begin{bmatrix} C_1 & C_3 \\ C_3^T & C_2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}^T \begin{bmatrix} x \\ y \end{bmatrix},$$

$$\text{s.t.} \quad Ax + By \leq b,$$

$$x, y \geq 0,$$

and the lower level problem in the variable y,

$$\text{(LLP}(x)) \quad \min_{y} \quad 1/2 y^T Q y + y^T S x + d^T y,$$

$$\text{s.t.} \quad By \leq b - Ax,$$

$$y \geq 0,$$

are strictly convex quadratic programming problems. We also assume that the constrained set

$$D(x) = \{y \in \mathbb{R}^m \colon By \leq b - Ax, y \geq 0\}$$

is nonempty for some values of x. As a consequence of this, problem LLP(x) has unique solutions for different values of x and the problem QBP has a global optimum (Ref. 7).

The following definitions characterize the feasible set of problem QBP as well as its extreme points.

Definition 2.1. The set $\{(x, y) \colon x \geq 0, y$ is optimal for LLP(x)$\}$ is called the induced region (IR) of problem QBP.

Since the lower level is a convex program in the y variable, the induced region is defined by the following linear complementarity conditions:

$$Qy + Sx + d + B^T\gamma - \beta = 0, \tag{1}$$

$$Ax + By + \alpha = b, \tag{2}$$

$$x, y, \alpha, \beta, \gamma \geq 0, \tag{3}$$

$$\alpha^T\gamma = y^T\beta = 0, \tag{4}$$

where $\alpha, \gamma \in \mathbb{R}^l$ and $\beta \in \mathbb{R}^m$. This result follows directly by replacing the lower level problem LLP(x) by its Karush–Kuhn–Tucker (KKT) conditions, which are sufficient by the convexity assumption.

Definition 2.2. $u = (x, y)$ is an extreme induced region (EIR) point if there exist α, β, γ such that $(x, y, \alpha, \beta, \gamma)$ is an extreme point of the polyhedral set defined by (1)–(3) and satisfies the complementarity conditions (4). An EIR point is said to be nondegenerate if the values of the basic variables are all positive; otherwise, it is called degenerate.

We assume throughout this paper that all EIR points are nondegenerate.

As in linear programming, two EIR points are said to be adjacent if their bases differ in exactly one column. It follows from these last definitions that a movement between two adjacent EIR points can be done by performing a pivot step that maintains the complementarity conditions. This can simply be done by not allowing two complementary variables to be simultaneously basic.

Definition 2.3 establishes a class of directions that plays an important role on the development of our approach to deal with quadratic bilevel programs.

Definition 2.3. d is an extreme induced (EIR) direction if it connects two adjacent EIR points.

In an EIR point is not a local minimum of the QBP problem, then there exists at least a descent EIR direction emanating from it. This result is stated in the next theorem and is used later to design a descent EIR point algorithm.

Theorem 2.1. Let u be an EIR point. If u is not a local minimum of problem QBP, then there is at least one descent EIR direction at u.

Proof. The nonoptimality at u implies the existence of at least one feasible descent IR direction d. From the piecewise linear property of the induced region (Ref. 6), such direction may be written as

$$d = \sum_{i=1}^{p} \mu_i d_i,$$

where

$$\sum_{i=1}^{p} \mu_i = 1, \qquad \mu_i > 0,$$

and d_i are EIR directions, $i = 1, \ldots, p$. If all directions d_i, $i, \ldots, p$, satisfy

$$\nabla F(u)^T d_i \geq 0,$$

then

$$\sum_{i=1}^{p} \mu_i \nabla F(u)^T d_i \geq 0.$$

By the convex linear combination of the direction d, this last condition implies

$$\nabla F(u)^T d \geq 0,$$

which contradicts the fact that d is a descent direction. Therefore, at least one of the EIR directions $d_1, \ldots, d_p$ is a descent direction. □

3. Descent Extreme Induced Region Point Algorithm

If $(\bar{x}, \bar{y})$ is a nondegenerate EIR point, then one of the following situations may occur:

(i) $(\bar{x}, \bar{y})$ is a local minimum for the QBP problem;

(ii) $(\bar{x}, \bar{y})$ is not a local minimum and there exists an EIR direction such that the corresponding adjacent EIR point $(\bar{\bar{x}}, \bar{\bar{y}})$ satisfies

$$F(\bar{\bar{x}}, \bar{\bar{y}}) < F(\bar{x}, \bar{y});$$

(iii) $(\bar{x}, \bar{y})$ is not a local minimum for the QBP problem and

$$F(\bar{\bar{x}}, \bar{\bar{y}}) \geq F(\bar{x}, \bar{y}),$$

for all adjacent EIR points $(\bar{\bar{x}}, \bar{\bar{y}})$; following the terminology used in Ref. 21, we call $(\bar{x}, \bar{y})$ a local star induced region (LSIR) point.

It is important to note that, if an LSIR point $(\bar{x}, \bar{y})$ is not a local minimum of problem QBP, then by Theorem 2.1 there exists at least a descent EIR direction at $(\bar{x}, \bar{y})$. To illustrate this situation, consider the following three-variable QBP problem:

$$\min_{x_1, x_2, y} \quad 1/2(x_1 - 4/5)^2 + 1/2(x_2 - 1/5)^2 + 1/2(y - 1)^2,$$

$$\text{s.t.} \quad 0 \leq x_1, x_2 \leq 1,$$

$$y \in \operatorname{argmin}\{(1/2)y^2 - y - x_1 y + 2x_2 y : 0 \leq y \leq 1\}.$$

The induced region for this simple QBP problem is the union of the following sets;

$$\{(x_1, x_2, y) \in \mathbb{R}^3 : x_1 \leq 1, x_2 \geq 0, -x_1 + 2x_2 \leq 0, y = 1\},$$

$$\{(x_1, x_2, y) \in \mathbb{R}^3 : -x_1 + 2x_2 + y = 1, 0 \leq y \leq 1\},$$

$$\{(x_1, x_2, y) \in \mathbb{R}^3 : x_1 \geq 0, x_2 \leq 1, -x_1 + 2x_2 \geq 1, y = 0\}.$$

The first of these sets is diagrammed in Fig. 1 and consists of the triangle with vertices $V_1 = (1, 0, 1)$, $V_2 = (0, 0, 1)$, $V_3 = (1, 1/2, 1)$. Although at the EIR point V_1, $\overrightarrow{V_1 V_2}$ and $\overrightarrow{V_1 V_3}$ are descent EIR directions, the values for the upper level objective function at the adjacent EIR points V_2 and V_3 are greater than the value at V_1. Hence, V_1 is an LSIR point.

It is easy to design an algorithm that finds at least an LSIR point for a quadratic bilevel program. The procedure starts by finding an initial EIR point. In each iteration, either the current EIR point is a LSIR point, or is a local minimum and the algorithm terminates, or an adjacent EIR point is obtained with a lower value of the upper level function. The procedure is then repeated. The steps of the algorithm are as follows.

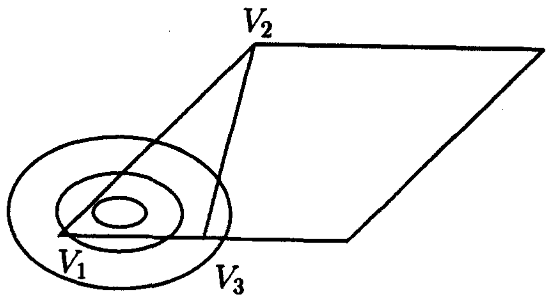

Fig. 1. LSIR point with two descent EIR directions.

Initial Step. Compute an initial EIR point u_0. Set $k = 0$.

General Step. Let D_k be the set of descent EIR directions at u_k,

$$D_k = \{d: \nabla F(u_k)^T d < 0 \text{ and } d \text{ is an EIR direction}\}.$$

(i) If $D_k \neq \varnothing$, select $d_k \in D_k$ such that

$$F(u_{k+1}) < F(u_k),$$

where u_{k+1} is the adjacent EIR point that is connected with u_k by the direction d_k. Set $k = k + 1$, and repeat this step. If such direction does not exist, stop: u_k is an LSIR point of problem QBP.

(ii) If $D_k = \varnothing$, stop: u_k is a local minimum for problem QBP.

As stated in the previous section, each iteration of the algorithm consists of a pivot step that maintains the complementarity condition. Hence, the computational effort of the algorithm is not too high.

The algorithm always terminates with an LSIR point u_k of problem QBP, provided all the EIR points are nondegenerate. However, only the case $D_k = \varnothing$ assures that u_k is a local minimum for problem QBP. This drawback of the algorithm motivates the use of another procedure such as the steepest descent method (Ref. 19). This is discussed in the next section.

Another important issue of the algorithm is the computation of an initial EIR point. Since the relaxed problem RP is a convex quadratic program and its constraint set is nonempty, then an optimal solution (x_R, y_R) exists and can be found in polynomial time (Ref. 22).

If (x_R, y_R) belongs to the induced region, then it is the global minimum of problem QBP. In general, such a situation does not occur. A first point of the induced region can be found by fixing $x = x_R$ and solving the lower level quadratic program LLP(x_R). Since Q is a symmetric positive-definite matrix and $D(x_R) \neq \emptyset$, this program has a unique solution $\bar{y}_R$, that can be found in polynomial time.

So, we can find in polynomial time an initial point of the induced region. However, such a solution is not in general an EIR point, since it does not correspond to a basic solution of the system of linear constraints defined by (1), (2), (3). An algorithm described in Ref. 23 can used to generate a basic feasible solution for the linear constraints. Such procedure has been proved to be polynomial (Ref. 24). Since the number of positive variables is reduced in each iteration, then the algorithm only visits points of the induced region (complementarity conditions are satisfied). Hence, we can find an initial EIR point in polynomial time.

As stated before, the algorithm terminates with an LSIR point of problem QBP provided all the EIR points are nondegenerate. If no such assumption is assumed, then checking that a given point is a local star minimum is a much more involved problem, since it might require the performance of a number of dual pivot steps to analyze all the extreme directions at the given point.

4. Modified Steepest-Descent Approach

In this section, we discuss the use of the steepest descent algorithm introduced in Ref. 19 for the solution of problem QBP. As in Ref. 19, we assume that the gradients of the active constraints at each point used by the algorithm are linearly independent.

In a given iteration k of this algorithm, the steepest descent direction $d_k = (z_k, w_k)$ at an induced region point $u_k = (x_k, y_k)$ is found by solving the following linear quadratic bilevel program (Ref. 19):

$$
\begin{aligned}
(\mathrm{LQBP}_k) \quad & \min_{z,w} \quad (C_1 x_k + C_3 y_k + c_1)^T z + (C_3^T x_k + C_2 y_k + c_2)^T w, \\
& \text{s.t.} \quad -1 \leq z_i \leq 1, \; i = 1, \ldots, n, \\
& \qquad w \in \operatorname{argmin}\{w^T Q w + 2 w^T S z: \\
& \qquad\qquad A'z + B'w \leq 0, \\
& \qquad\qquad (-\phi_k^T A')^T z + (Q y_k + S x_k + d)^T w = 0, \\
& \qquad\qquad w' \geq 0\},
\end{aligned}
$$

where $z \in \mathbb{R}^n$ and $w \in \mathbb{R}^m$. The matrices A' and B' contain all the rows of A and B corresponding to the active constraints at u_k. Similarly, the vector w'

is a subvector of w where only the indices i corresponding to zero variables $(y_k)_i$ are considered. Furthermore, ϕ_k are the multipliers associated to the active contraints at u_k.

If the optimal value of problem LQBP_k is greater than or equal to zero, then u_k is a local minimum of problem QBP. Otherwise, the optimal solution of problem LQBP_k is the steepest descent direction (z_k, w_k) and a new induced region point is found by

$$(x_{k+1}, y_{k+1}) = (x_k, y_k) + \sigma_k(z_k, w_k),$$

where σ_k is an appropriate stepsize.

If just a descent direction is required, there is no need to solve problem LQBP_k until the end. All the values between zero and the negative optimal value of problem LQBP_k correspond to descent directions (Ref. 19). This property should be exploited in the choice of the algorithm for the solution of problem LQBP_k. In this section, we show that the sequential LCP (Ref. 20) is quite suitable for this purpose. We also describe how an exact stepsize can be computed in an efficient way. These two procedures make the steepest descent algorithm much more attractive.

4.1. Use of Sequential LCP Method to Solve Problem LQBP_k. Since the lower level problem of LQBP_k is a convex program in w, it can be replaced by its KKT conditions. Thus, problem LQBP_k is equivalent to the following minimum linear complementarity problem:

$$\min_{z,w} \quad (C_1x_k + C_3y_k + c_1)^Tz + (C_3^Tx_k + C_2y_k + c_2)^Tw,$$

$$\text{s.t.} \quad 2Qw + 2Sz + B'^T\gamma' - \beta' + (Qy_k + Sx_k + d)^T\xi = 0,$$

$$A'z + B'w + \alpha' = 0,$$

$$(-\phi_k^TA')^Tz + (Qy_k + Sx_k + d)^Tw = 0,$$

$$\alpha'^T\gamma' = \beta'^Tw' = 0,$$

$$w', \alpha', \beta', \gamma' \geq 0, \qquad -1 \leq z_i \leq 1, \quad i = 1, \ldots, n,$$

where the dimension of the vectors α' and γ' is equal to the number of rows of B', and β' has the same dimension as w'.

The sequential LCP method (Ref. 20) looks for a global minimum of this last problem by solving a sequence of $\text{LCP}(\lambda_i)$. Each such problem can be posed as

$$(\mathrm{LCP}(\lambda_i)) \quad (C_1 x_k + C_3 y_k + c_1)^T z + (C_3^T x_k + C_2 y_k + c_2)^T w \leq \lambda_i,$$
$$2Qw + 2Sz + B'^T \gamma' - \beta' + (Qy_k + Sx_k + d)^T \xi = 0,$$
$$A'z + B'w + \alpha' = 0,$$
$$(-\phi_k^T A')^T z + (Qy_k + Sx_k + d)^T w = 0,$$
$$\alpha'^T \gamma' = \beta'^T w' = 0,$$
$$w', \alpha', \beta', \gamma' \geq 0, \qquad -1 \leq z_i \leq 1, \quad i = 1, \ldots, n,$$

where $\{\lambda_i\}$ is a strictly decreasing sequence.

The method stops when a LCP(λ_j) without a solution is found. In this case, the solution of the previous LCP(λ_{j-1}) is an ϵ-global solution for problem LQBP$_k$. The method works well to achieve the ϵ-global solution, but faces difficulties in establishing that such a solution has been found (Ref. 20). In fact, showing that a linear complementarity problem has no solution is a much harder task than just finding a solution to this problem.

As stated before, it is not required to find a global optimum for the LQBP$_k$. Instead, any solution (z, w) of problem LCP(λ_i) with $\lambda_i < 0$ is a descent direction. Since the sequential LCP algorithm solves a sequence of LCP(λ_i) with strictly decreasing values of λ_i, then it may terminate whenever a solution of problem LCP(λ_i) is found such that $\lambda_i < 0$. This overcomes the main drawback of the sequential LCP algorithm.

Now, suppose that a local minimum u_k is at hand. Then, there exist no descent directions emanating from u_k, which means that the optimal value of the LQBP$_k$ is nonnegative. So the sequential LCP has to perform its last step to assure that the local minimum has effectively been achieved.

As a final conclusion of this discussion, we can claim that the sequential LCP seems to be quite suitable to find a descent direction for the modified steepest descent algorithm. However, it is difficult to establish that the local minimum has been attained. It is known that checking local optimality in nonconvex quadratic programming is NP-hard (Ref. 25). The preceding discussion seems to indicate that the same property holds for quadratic bilevel programming. In Section 5, we show that checking local optimality in linear bilevel programming is NP-hard. Since linear bilevel programming is a particular instance of quadratic bilevel programming, the latter problem also shares this property.

4.2. Exact Stepsizes. Given an induced region point $u_k = (x_k, y_k)$ and a feasible IR direction $d_k = (z_k, w_k)$, an efficient criterion should be developed to compute the largest feasible stepsize $\sigma_{\max}$. For polyhedral sets, such a criterion is usually given in terms of a minimum quotient rule. In the case of an induced region, the boundaries are only implicitly defined and the problem of finding $\sigma_{\max}$ is more complicated. Next, we describe an efficient procedure for such purpose.

Let η be the number of active constraints at $u_k + \sigma d_k$, where σ is a small positive number. If $\eta = 0$, no multipliers exist in the dual constraints of the KKT conditions at $u_k + \sigma d_k$, whence

$$Q(y_k + \sigma w_k) + S(x_k + \sigma z_k) + d = 0.$$

The feasible IR direction $d_k = (z_k, w_k)$ should verify these conditions, and consequently the stepsize $\sigma_{\max}$ is the largest value of σ such that

$$A(x_k + \sigma z_k) + B(y_k + \sigma w_k) \leq b,$$

$$x_k + \sigma z_k \geq 0, \qquad y_k + \sigma w_k \geq 0.$$

Therefore, a minimum quotient rule is sufficient to compute $\sigma_{\max}$ when $\eta = 0$. Consider now the case $\eta > 0$. Let

$$r_i^T x + s_i^T y = t_i, \qquad i = 1, \ldots, \eta, \tag{5}$$

be the η lower level active constraints at $u_k + \sigma d_k$. The dual constraints of the KKT conditions at $u_k + \sigma d_k$ can be written as follows:

$$Q(y_k + \sigma w_k) + S(x_k + \sigma z_k) + d + \delta_1 s_1 + \cdots + \delta_\eta s_\eta = 0, \tag{6}$$

where δ_i, $i = 1, \ldots, \eta$, are the corresponding nonnegative multipliers. Since Q is a nonsingular matrix, then this last condition implies

$$y_k + \sigma w_k = -Q^{-1}S(x_k + \sigma z_k) - Q^{-1}d - \delta_1 Q^{-1}s_1 - \cdots - \delta_\eta Q^{-1}s_\eta.$$

Replacing this expression of $y_k + \sigma w_k$ in the active constraints (5), the following linear system in the multipliers δ_i, $i = 1, \ldots, \eta$, is obtained:

$$Z\delta = q' + \sigma q'', \tag{7}$$

where

$$Z = \begin{bmatrix} -s_1^T Q^{-1} s_1 & \cdots & -s_1^T Q^{-1} s_\eta \\ \vdots & \vdots & \vdots \\ -s_\eta^T Q^{-1} s_1 & \cdots & -s_\eta^T Q^{-1} s_\eta \end{bmatrix}, \qquad \delta = \begin{bmatrix} \delta_1 \\ \vdots \\ \delta_\eta \end{bmatrix},$$

$$q' = \begin{bmatrix} t_1 - r_1^T x_k + s_1^T Q^{-1} S x_k + s_1^T Q^{-1} d \\ \vdots \\ t_\eta - r_\eta^T x_k + s_\eta^T Q^{-1} S x_k + s_\eta^T Q^{-1} d \end{bmatrix},$$

$$q'' = \begin{bmatrix} -r_1^T z_k + s_1 Q^{-1} S z_k \\ \vdots \\ -r_\eta^T z_k + s_\eta Q^{-1} S z_k \end{bmatrix}.$$

By solving the two $\eta \times \eta$ linear systems

$$Zv' = q', \qquad Zv'' = q'',$$

the following linear relationship among all the η multipliers and the σ parameter is obtained:

$$\delta_i = v_i' + v_i'' \sigma, \qquad i = 1, \ldots, \eta.$$

We can measure the total computational effort to compute the vectors v' and v''. Indeed we can state the following theorem.

Theorem 4.1. The total number of systems required to compute the largest feasible stepsize $\sigma_{\max}$ is $\eta + 2$, namely, η systems with the matrix Q and two systems with the matrix $Z = [z_{ij}]_{n \times n}$ where $z_{ij} = -s_i Q^{-1} s_j$.

Proof. After solving the η systems $Q\xi = s_i$, $i = 1, \ldots, \eta$, only inner products are needed to prepare the data for solving the system (7). Then, two systems with the matrix Z are required to compute the vectors v' and v''. □

Since the matrix Q is symmetric positive definite, then its Cholesky factorization can be computed, that is, $Q = LL^T$, where L is a lower triangular matrix with positive diagonal elements. It is important to remark that this factorization has to be computed only once during the whole steepest descent procedure. So in each iteration, it is necessary to solve 2η triangular systems and two systems with the matrix Z.

After finding the vectors v' and v'', the stepsize $\sigma_{\max}$ is the largest value of σ such that

$$A(x_k + \sigma z_k) + B(y_k + \sigma w_k) \le b,$$

$$x_k + \sigma z_k \ge 0, \qquad y_k + \sigma w_k \ge 0,$$

$$v_i' + v_i'' \sigma \ge 0, \qquad i = 1, \ldots, \eta.$$

Hence, $\sigma_{\max}$ can be computed by using a minimum quotient rule.

For computing the exact stepsize σ_k, we consider the function

$$G(\sigma) = 1/2(u_k + \sigma d_k)^T C(u_k + \sigma d_k) + c^T(u_k + \sigma d_k), \qquad \sigma > 0,$$

where $c^T = (c_1^T, c_2^T)$. Since C is a symmetric positive-definite matrix, then G is a strictly convex function. Hence, the unconstrained minimizer σ_k' can be computed by solving $G'(\sigma) = 0$. Thus,

$$\sigma_k' = -(c^T d_k + d_k^T C u_k)/(d_k^T C d_k), \tag{8}$$

and the exact stepsize σ_k is computed as follows:

$$\sigma_k = \begin{cases} \sigma'_k, & \text{if } 0 < \sigma'_k < \sigma_{\max}, \\ \sigma_{\max} & \text{otherwise.} \end{cases} \tag{9}$$

To illustrate the computation of the exact stepsize σ_k, consider the following QBP problem:

$$\min_{x_1, x_2, y} \quad 1/2(x_1 - 1)^2 + 1/2(x_2 - 2/5)^2 + 1/2(y - 4/5)^2,$$
$$\text{s.t.} \quad 0 \leq x_1, x_2 \leq 1,$$
$$y \in \operatorname{argmin}\{1/2y^2 - y - x_1 y + 3x_2 y : 0 \leq y \leq 1\}.$$

The induced region for this simple three-variable QBP problem is the union of the following sets:

$$\{(x_1, x_2, y) \in \mathbb{R}^3 : x_1 \leq 1, x_2 \geq 0, -x_1 + 3x_2 \leq 0, y = 1\},$$
$$\{(x_1, x_2, y) \in \mathbb{R}^3 : -x_1 + 3x_2 + y = 1, 0 \leq y \leq 1\},$$
$$\{(x_1, x_2, y) \in \mathbb{R}^3 : x_1 \geq 0, x_2 \leq 1, -x_1 + 3x_2 \geq 1, y = 0\}.$$

Figure 2 describes the first of these sets, the triangle of vertices $V_1 = (1, 0, 1)$, $V_2 = (0, 0, 1)$, and $V_3 = (3/2, 1/2, 1)$. If $u_0 \equiv V_1$, then the steepest descent direction d_0 is $(0, 1/3, 0)$ and a new induced region point $u_1 \equiv V_4$ is computed by $u_1 = u_0 + \sigma_0 d_0$, where $\sigma_0 = \sigma_{\max} = 1/3$. In the second iteration, $d_1 = (0, 1/15, -1/5)$ and $u_2 = u_1 + \sigma_1 d_1 = (1, 2/5, 4/5) \equiv V_5$, where $\sigma_1 = \sigma'_1 = \sqrt{10}/15$. The point u_2 is a local minimum, and the steepest descent algorithm terminates. Note that, in the first iteration, the stepsize

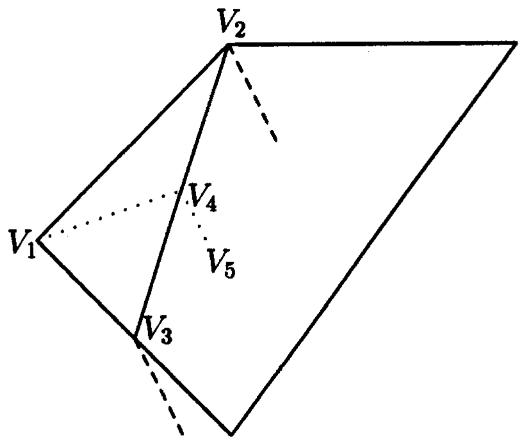

Fig. 2. Computation of exact stepsizes $\sigma_{\max}$ and σ'_k.

σ'_0 is not feasible and the displacement along the direction d_0 is made by the largest feasible stepsize $\sigma_{\max}$. This is not the case of the second iteration, where the stepsize σ'_1 is feasible.

It is easy to see that computing σ'_k is a much less involved task than finding the value of the stepsize $\sigma_{\max}$. An alternative procedure can be proposed which tries to avoid the computation of $\sigma_{\max}$ at the expense of solving a strictly convex program. The steps of this technique are presented below

(i) Compute σ'_k as stated in (8) and set $\bar{x} = x_k + \sigma'_k z_k$.

(ii) Solve, if possible, the strictly convex quadratic program LLP($\bar{x}$). Let $\bar{y}$ be the optimal solution of this problem. If $\bar{y} = y_k + \sigma'_k w_k$, then consider the new point

$$u_{k+1} = (x_{k+1}, y_{k+1}) = (\bar{x}, \bar{y}).$$

(iii) Otherwise σ'_k is not a local feasible stepsize and $\sigma_{\max}$ has to be computed. Furthermore,

$$u_{k+1} = u_k + \sigma_{\max} d_k.$$

Since the quadratic program LLP($\bar{x}$) can be solved in polynomial time in a quite small number of iterations that do not depend on the number of active constraints (Ref. 26), this procedure may be an alternative approach, particularly when the number of active constraints η is large. However, we may be forced to compute $\sigma_{\max}$ if σ'_k is not feasible.

4.3. Hybrid Approach. In the two previous sections, we have described two algorithms for the solution of the quadratic bilevel programming problem. The descent EIR point algorithm is quite simple, but cannot always assure a local minimum for the QBP problem. On the other hand, the steepest descent method is much more involved but always terminates with a local minimum of the QBP problem. So a hybrid approach can be designed that tries to exploit the advantages of these two procedures. The steps of the resulting algorithm are presented below.

Step 1. Apply the descent EIR point algorithm. If the algorithm terminates with a local minimum, stop. Otherwise, let u_k be the LSIR point obtained at the end of the procedure.

Step 2. Solve problem LQBP$_k$ to get a decent direction d_k. If the optimal value of this problem is greater than or equal to zero, stop: u_k is a local minimum of problem QBP. Otherwise, compute σ_k as in (9) and set $u_{k+1} = u_k + \sigma_k d_k$. Set $k = k + 1$, and repeat Step 2.

The possibility of moving from Step 2 to Step 1 after some iterations in Step 2 could also be considered. In this case, an initial EIR point can be found by using the procedure discussed in Section 3. This strategy could be used either to improve the convergence of the steepest descent algorithm or, after the termination of the algorithm, to start the process again to look for a better local minimum.

Consider again problem QBP and suppose that the upper level function F is concave, that is, C is negative semidefinite. Then, a well-known result for concave programming also holds for the concave QBP problem.

Theorem 4.2. If F is a concave (strictly concave) function, then at least one (every) local minimum of problem QBP is attained at an extreme induced region point.

Proof. To prove the theorem, we start by showing that all the induced region points in a given face form a polyhedral set. In fact, suppose that a given face is the set of η active constraints of the form (5). The dual constraints of the KKT conditions (6) represent a polyhedral set in $\mathbb{R}^{n+m+\eta}$. The projection of this polyhedral set onto $\mathbb{R}^{n+m}$ is also a polyhedral set. The intersection of this last convex set with the given face is a polyhedral IR subset of that face.

Now, let u be a given local minimum for the QBP problem. As we have proved before, the induced region of problem QBP is a finite union of polyhedral sets,

$$\mathrm{IR} = \bigcup_{i=0}^{K} P_i,$$

where P_i is a polyhedral set, $i = 1, \ldots, K$. Thus, there exists at least one $k \in \{1, \ldots, K\}$ such that $u \in P_k$. Since u is a local minimum over P_k, the result follows immediately from the theory concerning the minimization of concave functions. □

Suppose that we apply the descent algorithm described in Section 3 to problem QBP when F is a concave function. If the last EIR point is nondegenerate, then it is an LSIR point and, by Theorems 2.1 and 4.2, it is also a local minimum for problem QBP. So the modified steepest descent method is not required in this last case. However, if the last EIR point obtained by the descent EIR point algorithm is degenerate, then there is no guarantee that such point is a local minimum.

5. Checking Local Optimality in Bilevel Programming Is NP-Hard

In this section, we prove that checking strict local optimality and checking local optimality in linear bilevel programming are NP-hard problems. To prove these results, we use the same ideas described in Pardalos and Schnitger (Ref. 25) for nonconvex quadratic programming, where the problem of checking (strict or not) local optimality was proved to be equivalent to a 3-satisfiability (3-SAT) problem. A classical survey on intractability issues can be found in Ref. 27.

These last two results have been proved using a special indefinite program whose set of constraints,

$$A_S x \geq 3/2 + c, \tag{10a}$$

$$1/2 - x_0 \leq x_i \leq 1/2 + x_0, \qquad i = 1, \ldots, n, \tag{10b}$$

$$x_i \geq 0, \qquad i = 0, \ldots, n, \tag{10c}$$

is associated to a given instance S of the 3-SAT problem.

To prove our complexity results, we start by considering linear bilevel programs containing the constraints (10) as upper level constraints. Then we show that these problems satisfy some properties, similar to those presented in Ref. 25, that enable one to establish the desired conclusions.

Theorem 5.1. Checking strict local optimality in linear bilevel programming is NP-hard.

Proof. Consider the following instance of a linear bilevel program

$$\begin{aligned} \min_{x,l,m,z} \quad & F(x, l, m, z) = \sum_{i=1}^{n} z_i, \\ \text{s.t.} \quad & A_S x \geq 3/2 + c, \\ & 1/2 - x_0 \leq x_i \leq 1/2 + x_0, \qquad i = 1, \ldots, n, \\ & x_i \geq 0, \qquad i = 0, \ldots, n, \\ & l, m, z \in \operatorname{argmax}\left\{ \sum_{i=1}^{n} z_i : x_i - l_i = 1/2 - x_0, x_i + m_i = 1/2 + x_0, \right. \\ & \qquad \left. z_i \leq l_i, z_i \leq m_i, i = 1, \ldots, n, z \geq 0 \right\}. \end{aligned}$$

Since all the variables z_i are forced to be nonnegative, then

$$F(x, l, m, z) \geq 0.$$

Consider now the point x^* defined by

$$x_0^* = 0, \qquad x_i^* = 1/2, \qquad i = 1, \ldots, n. \tag{11}$$

Then, x^* satisfies the upper level constraints. Furthermore, $l^* = 0$, $m^* = 0$, and $z^* = 0$ is the optimal solution of the lower level problem when x is set equal x^*. Hence, $(x^*, l^*, m^*, z^*) = 0$, then (x^*, l^*, m^*, z^*) is a global minimum of the linear bilevel program.

As in Ref. 25, the theorem is proved by establishing that $F(x, l, m, z) = 0$ if and only if $x_i \in \{1/2 - x_0, 1/2 + x_0\}$, for all $i = 1, \ldots, n$. If this last condition holds, then $l_i = 0$ or $m_i = 0$ and $z_i = 0$ for all $i = 1, \ldots, n$. This implies that $F(x, l, m, z) = 0$. To show the converse, suppose that $x_i \neq 1/2 - x_0$ and $x_i \neq 1/2 + x_0$ for some i. Since the lower level program is a maximization problem, then z_i must be positive, which implies that $F(x, l, m, z) > 0$. □

Theorem 5.2. Checking local optimality in linear bilevel programming in NP-hard.

Proof. Consider the following linear bilevel program

$$\begin{aligned}
\min_{x,l,m,z,w} \quad & F(x, l, m, z, w) = \sum_{i=1}^{n} z_i - 1/2n \sum_{i=1}^{n} w_i, \\
\text{s.t.} \quad & A_S x \geq 3/2 + c, \\
& 1/2 - x_0 \leq x_i \leq 1/2 + x_0, \qquad i = 1, \ldots, n, \\
& x_i \geq 0, \qquad i = 0, \ldots, n, \\
& l, m, z, w \in \operatorname{argmax} \Big\{ \sum_{i=1}^{n} z_i - \sum_{l=1}^{n} w_i : x_i - l_i = 1/2 - x_0, \\
& \qquad x_i + m_i = 1/2 + x_0, \\
& \qquad z_i \leq l_i, z_i \leq m_i, i = 1, \ldots, n \\
& \qquad w_i \geq x_i - 1/2, \\
& \qquad w_i \geq 1/2 - x_i, i = 1, \ldots, n, z, w \geq 0 \Big\}.
\end{aligned}$$

Let IR be the induced region of this program. As before, consider the point x^* defined by (11). If we fix $x = x^*$ and solve the lower level program, we get

$$l_i^* = m_i^* = z_i^* = w_i^* = 0, \qquad i = 1, \ldots, n.$$

Hence, $(x^*, l^*, m^*, z^*, w^*) \in \text{IR}$. Furthermore,

$$F(x^*, l^*, m^*, z^*, w^*) = 0.$$

Therefore this point $(x^*, l^*, m^*, z^*, w^*)$ can play the same role as x^* in Theorem 2 of Ref. 25. So, we can prove our result if we are able to show that

$$F(x, l, m, z, m) > 0,$$

for all $(x, l, m, z, w) \in \mathrm{IR}$ satisfying $x_1 \geq 1/2 - x_0/3$ and $x_0 > 0$.

Since the lower level program is a maximization problem, then

$$z_i = \min\{l_i, m_i\}, \qquad i = 1, \ldots, n.$$

Hence,

$$\sum_{i=1}^{n} z_i = \sum_{i=1}^{n} \min\{l_i, m_i\} = \sum_{i=1}^{n} \min\{x_i - 1/2 + x_0, \ -x_i + 1/2 + x_0\}.$$

But

$$x_1 \geq 1/2 - x_0/3,$$

and this implies that

$$\sum_{i=1}^{n} z_i \geq \min\{x_1 - 1/2 + x_0, \ -x_1 + 1/2 + x_0\} = 2/3x_0. \tag{12}$$

On the other hand, since

$$w_i \geq x_i - 1/2, \quad w_i \geq 1/2 - x_i, \qquad i = 1, \ldots, n,$$

and the lower level program is a maximization problem, then we must have

$$w_i = |x_i - 1/2|, \qquad i = 1, \ldots, n.$$

Furthermore, the upper level constraints imply that

$$|x_i - 1/2| \leq x_0, \qquad i = 1, \ldots, n.$$

Hence,

$$1/2n \sum_{i=1}^{n} w_i \leq 1/2n \sum_{i=1}^{n} x_0 = x_0/2. \tag{13}$$

Therefore by (12) and (13), we get

$$F(x, l, m, z, w) \geq 2/3x_0 - x_0/2 = x_0/6 > 0,$$

and this proves the theorem. □

Since the upper level constraints do not contain the lower level variables z_i, l_i, m_i, w_i, they can be moved to the lower level problem. Hence, the theorems are also valid if the upper level constraints only contain bounds for the values of the upper level variables.

6. Conclusions

In this paper, a descent framework for quadratic bilevel programming has been discussed. Different strategies depending on the properties of the upper level function have been introduced. The rules for computing exact stepsizes proposed in Section 4.2 can be used by any descent method as long as the lower level problem has a strictly convex quadratic structure. We have also dealt with the intractability of quadratic bilevel programs, by proving that checking local optimality is a NP-hard problem.

The case where the upper level function F is strictly convex but not quadratic does not bring many changes to the descent approach described in the last sections. The sequential LCP method can still be applied and only the computation of the stepsize σ_k requires a different technique, since no exact procedures solve the equation $G'(\sigma) = 0$ directly. Theorem 4.2 also holds for general concave functions, whence the descent EIR point algorithm terminates with a local minimum provided all the EIR points are nondegenerate.

In this paper, we have not considered the possible occurrence of degeneracy in the descent algorithms. Degeneracy is an important issue and deserves some attention in the future. Another important area of future research is the extension of the descent techniques described in this paper to the solution of bilevel programs with linear or nonlinear upper level constraints in both the x and y variables. It is our opinion that the success of this research will play a major role in the development of techniques for finding global minima of nonlinear bilevel programs in which the lower level function is quadratic and strictly convex.

References

1. ANANDALINGAM, G., and FRIESZ, T., Editors, *Hierarchical Optimization*, Annals of Operations Research, Vol. 34, 1992.
2. BARD, J., and MOORE, J., *A Branch-and-Bound Algorithm for the Bilevel Programming Problem*, SIAM Journal on Scientific and Statistical Computing, Vol. 11, pp. 281–292, 1990.
3. HANSEN, P., JAUMARD, B., and SAVARD, G., *New Branching-and-Bounding Rules for Linear Bilevel Programming*, SIAM Journal on Statistical and Scientific Computing, Vol. 13, pp. 1194–1217, 1992.
4. JÚDICE, J., and FAUSTINO, A., *A Sequential LCP Method for Bilevel Linear Programming*, Annals of Operations Research, Vol. 34, pp. 89–106, 1992.
5. AL-KHAYYAL, F., HORST, R., and PARDALOS, P., *Global Optimization of Concave Functions Subject of Quadratic Constraints: An Application in Nonlinear Bilevel Programming*, Annals of Operations Research, Vol. 34, pp. 125–147, 1992.

6. BARD, J., *Convex Two-Level Programming*, Mathematical Programming, Vol. 40, pp. 15–27, 1988.
7. EDMUNDS, T., and BARD, J., *Algorithms for Nonlinear Bilevel Mathematical Programs*, IEEE Transactions on Systems, Man, and Cybernetics, Vol. 21, pp. 83–89, 1991.
8. JAUMARD, B., SAVARD, G., and XIONG, J., *An Exact Algorithm for Convex Bilevel Programming*, Optimization Days, Montreal, Canada, 1992.
9. AIYOSHI, E., and SHIMIZU, K., *Hierarchical Decentralized System and Its New Solution by a Barrier Method*, IEEE Transactions on Systems, Man, and Cybernetics, Vol. 11, pp. 444–449, 1981.
10. AIYOSHI, E., and SHIMIZU, K., *A Solution Method for the Static Constrained Stackelberg Problem via Penalty Method*, IEEE Transactions on Automatic Control, Vol. 29, pp. 1111–1114, 1984.
11. BI, Z., CALAMAI, P., and CONN, A., *An Exact Penalty Function Approach for the Nonlinear Bilevel Programming Problem*, Technical Report 180-O-170591, University of Waterloo, 1991.
12. ISHIZUKA, Y., and AIYOSHI, E., *Double Penalty Method for Bilevel Linear Programming*, Annals of Operations Research, Vol. 34, pp. 73–88, 1992.
13. FLORIAN, M., and CHEN, Y., *A Bilevel Programming Approach to Estimating O-D Matrix by Traffic Counts*, Report CRT-750, Centre de Recherche sur les Transports, 1991.
14. KOLSTAD, C., and LASDON, L., *Derivative Evaluation and Computational Experience with Large Bilevel Mathematical Programs*, Journal of Optimization Theory and Applications, Vol. 65, pp. 485–499, 1990.
15. BARD, J., *Optimality Conditions for the Bilevel Programming Problem*, Naval Research Logistics Quarterly, Vol. 31, pp. 13–26, 1984.
16. CLARKE, P., and WESTERBERG, A., *A Note of the Optimality Conditions for the Bilevel Programming Problem*, Naval Research Logistics, Vol. 35, pp. 413–418, 1988.
17. CHEN, Y., and FLORIAN, M., *The Nonlinear Bilevel Programming Problem: A General Formulation and Optimality Conditions*, Report CRT-794, Centre de Recherche sur les Transports, 1991.
18. DEMPE, S., *A Necessary and a Sufficient Optimality Condition for Bilevel Programming Problems*, Optimization (to appear).
19. GAUVIN, J., and SAVARD, G., *The Steepest-Descent Method for the Nonlinear Bilevel Programming Problem*, Report G-90-37, GERAD (Groupe d'Études et de Recherche en Analyse des Décisions), 1990.
20. JÚDICE, J., and FAUSTINO, A., *The Linear-Quadratic Bilevel Programming Problem*, INFOR (to appear).
21. AL-KHAYYAL, F., *An Implicit Enumeration Procedure for the General Linear Complementarity Problem*, Mathematical Programming Studies, Vol. 31, pp. 1–20, 1987.
22. KOJIMA, M., MIZUNO, S., and YOSHISE, A., *A Polynomial-Time Algorithm for a Class of Linear Complementarity Problems*, Mathematical Programming, Vol. 50, pp. 331–342, 1991.

23. MURTY, K., *Linear Programming*, John Wiley and Sons, New York, 1983.
24. MEGIDDO, N., *On Finding Primal and Dual Optimal Bases*, ORSA Journal on Computing, Vol. 3, pp. 63–65, 1991.
25. PARDALOS, P., and SCHNITGER, G., *Checking Local Optimality in Constrained Quadratic Programming is NP-Hard*, Operations Research Letters, Vol. 7, pp. 33–35, 1988.
26. CARPENTER, T., LUSTIG, I., MULVEY, J., and SHANNO, D., *Higher-Order Predictor-Corrector Interior-Point Methods with Application to Quadratic Objectives*, RUTCOR Research Report RRR 67-90, Rutgers University, 1990.
27. GAREY, M., and JOHNSON, D., *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W. H. Freeman and Company, New York, New York, 1979.