On the Convergence of the Algorithm for Bilevel Programming Problems by Clegg and Smith

Guy Cohen, Jean-Pierre Quadrat, Laura Wynter

▶ **To cite this version:**

Guy Cohen, Jean-Pierre Quadrat, Laura Wynter. On the Convergence of the Algorithm for Bilevel Programming Problems by Clegg and Smith. [Research Report] RR-4185, INRIA. 2001. inria-00072438

HAL Id: inria-00072438
https://inria.hal.science/inria-00072438

Submitted on 24 May 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

On the convergence of the algorithm for bilevel programming problems by Clegg and Smith

Guy Cohen, Jean-Pierre Quadrat, and Laura Wynter

N° 4185

May, 2001

THÈME 4

ISSN 0249-6399 ISRN INRIA/RR--4185--FR+ENG

INRIA
ROCQUENCOURT

On the convergence of the algorithm for bilevel programming problems by Clegg and Smith

Guy Cohen,* Jean-Pierre Quadrat,† and Laura Wynter‡

Thème 4 — Simulation et optimisation
de systèmes complexes

Projet METALAU

Rapport de recherche n° 4185 — May, 2001 — 8 pages

Abstract: In [1, 2], Smith and colleagues present an algorithm for solving the bilevel programming problem. We show that the points reached by the algorithm are not stationary points of bilevel programs in general. We further show that, with a minor modification, this method can be expressed as an inexact penalty method for gap function-constrained bilevel programs.

Key-words: bilevel optimization, hierarchical optimization, Stackelberg game, optimality conditions

* École Nationale des Ponts et Chaussées, 6-8, avenue Blaise Pascal, Cité Descartes, Champs-sur-Marne, 77455 Marne-la-Vallée cedex 2, France
† INRIA Rocquencourt
‡ INRIA Rocquencourt and PRISM, Université de Versailles

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau, Rocquencourt, BP 105, 78153 Le Chesnay Cedex (France)
Téléphone : +33 1 39 63 55 11 — Télécopie : +33 1 39 63 53 30

Sur la convergence de l'algorithme pour l'optimisation biniveaux de Clegg et Smith

Résumé : Dans les références [1, 2], Smith et collègues présentent un algorithme pour résoudre le problème de l'optimisation bi-niveaux. Nous montrons que les points obtenus par l'algorithme ne sont pas de points stationnaires des problèmes bi-niveaux en général. On montre également qu'avec une modification, cette méthode peut être exprimée comme un algorithme de pénalité inexacte pour des problèmes bi-niveaux d'optimisation, dans lesquels le problème au second niveau est défini avec une fonction de mérite, tel que l'on utilise pour transformer les inéquations variationnelles en problème d'optimisation.

Mots-clés : programmes d'optimisation biniveaux, optimisation hiérarchique, jeux de Stackelberg, conditions d'optimalité

1 Introduction

The bilevel programming problem arises in many contexts of transportation planning. The network design problem, optimal pricing problem, inverse O-D matrix estimation, and the optimal signal setting problem are just a few of the important applications of bilevel programming to transportation. The problem can be expressed as follows: the leader, or system manager, wishes to determine an optimal policy as a function of his or her control variables, $x \in R^n$, and of the users' reponses to these controls, $y \in R^m$, where the user response generally takes the form of a network flow, that is $y = \{y_i\}$, for $i \in A$, where A is the set of arcs on the network. The system manager then seeks to minimize a function of both x and y, where some constraints may be imposed upon x as well as the fact that y should be an user equilibrium flow, parametrized by the control vector, x. The problem is known as a bilevel programming problem since the system manager's optimization problem encompasses the user equilibrium problem, written as a variational inequality, which needs to be solved in order to evaluate the system manager's cost function. It may be expressed as follows:

$$\min f(x, y) \tag{1}$$

subject to

$$x \in X, \tag{2}$$
$$y \in S(x), \tag{3}$$

where $S(x) := \{y \mid T(x,y)^T(y-z) \leq 0, z \in Y(x)\}$, and $x \in R^n, y \in R^m, f : R^{n+m} \to R, T : R^{n+m} \to R^m$, $Y(x) := \{y \mid h(x,y) \leq 0\}, h : R^{n+m} \to R^p$. The implicit function, $S(x)$, can be rewritten as the solution set of an equivalent optimization problem, via a gap function for variational inequality problems. That is,

$$S(x) := \{y \mid y \in \arg\min_y g(x,y), y \in Y(x)\}. \tag{4}$$

The gap function proposed by Smith [6] for variational inequality problems defined over polyhedral sets is

$$g(y) := \sum_{i \in I} ([T(y)^T(y - z_i)]_+)^p \tag{5}$$

where $y_i, i \in I$, are the extreme points of the polyhedral feasible set, $[\cdot]_+ := \max\{0, \cdot\}$, and $p > 1$. Other gap functions for variational inequality problems can be found in [5]

In [1] and [2], a descent algorithm called the halfplane method was proposed for solving the bilevel programming problem. In these references, descent is sought simultaneously in both the upper level and lower level variables, x, and y without satisfying the equilibrium conditions during the course of the algorithm. This is different from methods that solve the equilibrium problem explicitly to obtain sensitivity information, or implictly by obtaining directional derivatives of the implicit function, $S(x)$, or rewriting the VIP as a system of equalities and inequalities, and solving a nonlinearly constrained mathematical program. In particular, Clegg and Smith propose updating simultaneously both x and y according to the following direction, $d \in R^{n \times m}$:

$$d := -\nabla_w g(w) + P_H[-\nabla_w f(w)], \tag{6}$$

where

$$H := \{\delta \mid \delta^T \nabla_w g(w) \leq 0\},$$

and $w := (x, y)$.

A similar algorithm, entitled the cone method, was proposed as well in [1], the sole difference between the two being that the set H is divided into two subsets and the projection made over the intersection of the two subsets. The use of one or the other definitions does not alter the results that follow in the remainder of this note.

We note that the definitions above assume that the upper and lower level problems in f and g, respectively, are unconstrained. This requires a penalization of the constraints, if present.

2 Optimality conditions for bilevel programming problems

Suppose that the function $T(x, y)$ is strictly monotonic in y. Then, the response function $y(x)$ is unique. If, furthermore, the lower-level problem is unconstrained, that is

$$S(x) := \{y \mid y \in \arg\min g(x, y), y \in R^m\},$$

where $g(x, y)$ is a gap function as described above, for the mapping $T(x, y)$, and T is differentiable in its first and second arguments, then $y(x)$ is differentiable.

Suppose now that the upper level problem is unconstrained in x, that is $\mathcal{X} = \Re^n$. In this case, it is straightforward to write the optimality conditions for the bilevel programming problem (1). Defining $\phi : x \mapsto f(x, y(x))$, we can express the total derivative of ϕ as follows:

$$\nabla\phi = \nabla_x f + \nabla y(x) \nabla_y f, \tag{7}$$

where $y(x)$ is the unique solution to the lower level problem. Making use of the gap function (5), it is possible to express $\nabla y(x)$ as the solution to

$$\nabla_y g(x, y^*(x)) = 0.$$

Differentiating this equality, we obtain an explicit expression for $\nabla y(x)$:

$$\nabla^2_{yx} g + \nabla y(x) \nabla^2_{yy} g = 0,$$

so that, when $\nabla^2_{yy} g$ is nonsingular,

$$\nabla y(x) = -\nabla^2_{yx} g [\nabla^2_{yy} g]^{-1}. \tag{8}$$

Then, substituting (8) into (7), we obtain

$$\nabla\phi = \nabla_x f - \nabla^2_{yx} g [\nabla^2_{yy} g]^{-1} \nabla_y f. \tag{9}$$

An analogous formula to (8) exists for the case in which the lower level problem is constrained. Then, we may write the Lagrangian of the lower level problem as

$$L(x, y, \pi) := g(x, y) + \pi^T h(x, y),$$

where $\pi \in R^p$ is the vector of Lagrange multipliers. Following [4], under the assumptions of linear independence of the binding constraints and strict complementarity, as well as the strict monotonicity of the lower level mapping $T(x, \cdot)$, the optimal response function $y(x)$ is still differentiable; its derivative is given below:

$$\begin{aligned} \nabla y(x) \quad = \quad & -\nabla^2_{yy} L^{-1} (I - \nabla_y h_I^T (\nabla_y h_I \nabla^2_{yy} L^{-1} \nabla_y h_I^T)^{-1} \nabla_y h_I \nabla^2_{yy} L^{-1} \nabla^2_{yx} L \\ & -\nabla^2_{yy} L^{-1} \nabla_y h_I^T (\nabla_y h_I \nabla^2_{yy} L^{-1} \nabla_y h_I^T)^{-1} \nabla_x h_I, \end{aligned}$$

where all derivatives are evaluated at the stationary point (x^*, y^*, π^*), and I represents the set of active constraints indices, that is $I := \{i | h_i(x, y) = 0\}$. In [4], it is further shown that the directional derivative in a direction s, expressed as $y'(x^*; s) = \nabla y(x^*)^T s$, is given by the solution to a particular quadratic program subject to linear equality constraints, where the constraints depend upon the set of active constraints of the lower level problem at that point. (Analogously, then, the second term of (7) can be seen to be equivalent to the directional derivative of $y(x)$ in the direction $\nabla_y f$ when the lower level problem is unconstrained.)

When strict complementarity is not satisfied, as is often the case with the traffic equilibrium problem, the optimal response function $y(x)$ is no longer differentiable. However, under the other assumptions imposed, $y(x)$ is still directionally differentiable. In this case, [4] show that the directional derivative $y'(x; s)$ in direction s is the solution to a nearly identical quadratic program as described above, the difference being that active constraints with zero multiplier values need only be satisfied up to an inequality, rather than with equality.

3 Stationarity of the algorithm of Clegg and Smith

Given the simple form of the projection, it is possible to write explicitly the formula for the direction d of (6).

THEOREM 1 *The direction d of (6) is equivalent to the following expression:*

$$d = (-1 - \pi) \nabla g(x, y) - \nabla f(x, y), \tag{10}$$

where

$$\pi = \max\{0, -\frac{\nabla f^T \nabla g}{\nabla g^T \nabla g}\}. \tag{11}$$

Proof: Since $P_H[-\nabla f(x, y)]$ is precisely the solution $\bar{z} \in R^{n+m}$ to

$$\bar{z} := \arg\min_z \frac{1}{2} \|z + \nabla f(x, y)\|^2 \tag{12}$$

such that

$$z^T \nabla g(x,y) \leq 0, \tag{13}$$

we may express the direction d of (6) as $d = -\nabla g + \bar{z}$.

The Lagrangian of the projection problem of (12)–(13) is given by the following, where all gradients are evaluated at (x,y):

$$L(z,\pi) = \frac{1}{2}\|z + \nabla f\|^2 + \pi z^T \nabla g. \tag{14}$$

Taking the partial derivative with respect to z we obtain

$$L'_z = z + \nabla f + \pi \nabla g = 0, \tag{15}$$

and therefore

$$z = -\nabla f - \pi \nabla g. \tag{16}$$

Then, substituting this value of z back into the Lagragian formula of (14),we have

$$L(\pi) = \frac{1}{2}\|-\pi \nabla g\|^2 + \pi(-\nabla f - \pi \nabla g)^T \nabla g. \tag{17}$$

Simplifying and taking the derivative of $L(\pi)$, we obtain

$$L'_\pi = \pi \nabla g^T \nabla g - \nabla f^T \nabla g - 2\pi \nabla g^T \nabla g. \tag{18}$$

Solving for π and taking into account complementarity requirements, we have that

$$\pi = \max\{0, -\frac{\nabla f^T \nabla g}{\nabla g^T \nabla g}\}. \tag{19}$$

Substituting this value of π into the Lagrangian $L(z,\pi)$, taking derivatives with respect to and solving for z, we have that

$$\overline{z} = -\nabla f - \pi \nabla g, \tag{20}$$

which allows us to write the explicit formula for the direction as in (10),

$$d = (-1-\pi)\nabla g - \nabla f, \quad \text{where} \quad \pi = \max\{0, -\frac{\nabla f^T \nabla g}{\nabla g^T \nabla g}\}, \tag{21}$$

which concludes the proof. □

Using the explicit representation of d from Theorem 1, it is straightforward to show that the stationarity conditions of the algorithm do not correspond to the optimality conditions of bilevel programs.

Proposition 2 *The set of points reached by the algorithm of [1, 2] do not coincide with the set of stationary points of the bilevel programming problem (1).*

Proof: We proceed by contradiction. Suppose that every limit point (x^*, y^*) of $\{(x^k, y^k)\}$ produced by the algorithm is a stationary point for the bilevel programming problem. Then, when $d^k = 0$, for some k, the corresponding (x^k, y^k) must be stationary for (1).

Rewriting the expression for d from Theorem 1, and using the fact that the lower-level problem is unconstrained, we have that $\nabla_y g(x, y) = 0$ at optimality, which gives:

$$d = \left(-1 + \min\{0, \frac{\nabla f^T \nabla g}{\nabla g^T \nabla g}\right) \begin{pmatrix} \nabla_x g \\ 0 \end{pmatrix} - \begin{pmatrix} \nabla_x f \\ \nabla_y f \end{pmatrix}. \tag{22}$$

The use of a gap function, g for the lower-level variational inequality has the property that $\nabla_y g(x, y) = 0 \Rightarrow \nabla_x g(x, y) = 0$. Indeed, $g(x, y(x)) = 0$ by construction of the gap function. This reduces the system (22) to the following:

$$d = -\begin{pmatrix} \nabla_x f \\ \nabla_y f \end{pmatrix}. \tag{23}$$

However, for

$$d = \begin{pmatrix} d_x \\ d_y \end{pmatrix} \tag{24}$$

to be zero, it would be necessary for $\nabla_x f(x, y) = \nabla_y f(x, y) = 0$, while stationarity for the bilevel programming problem (1) requires that

$$\nabla_x f - \nabla^2_{yx} g [\nabla^2_{yy} g]^{-1} \nabla_y f = 0. \tag{25}$$

It is clear that (23) need not hold for (25) to be satisfied, and that therefore the set of points satisfying (22) is not equivalent to the set of stationary points for bilevel programs. □

4 Relation to penalty methods for bilevel programs

Penalty methods for gap function-constrained bilevel programs were summarized in Marcotte and Zhu (1996). In that reference, the authors show in which cases the penalty function is exact and inexact, and provide general convergence results for algorithms which penalize the gap function reformulation of the lower-level variational inequality of a bilevel program, and iterate jointly in both upper and lower-level variables.

From the systems (22) and (23), one can see how to convert the algorithm of Clegg and Smith to the classic method of penalizing the gap function of the bilevel program. Proposition 1 states that d should be identically zero at the optimum. The first term of (6) vanishes at the optimum by construction of the gap function, as is shown in Proposition 2.

Then, by multiplying the second term of (6) by a parameter $1/\epsilon^k$ such that $\lim_{k\to\infty} \epsilon^k = \infty$, we see that the second term would also be forced to zero, as is required by the optimality conditions for bilevel programs.

We see, further, that the algorithm of [1] and [2] is then rendered equivalent to an inexact penalty method, and the proof of convergence then follows from [3].

References

[1] J. Clegg and M.J. Smith. Cone projection versus half-space projection for the bilevel optimisation of transportation networks. Transportation Research B, 35, 71–82 (2001).

[2] A. Battye, A. Clune, M. Smith, and Y. Xiang. Cone fields and the cone projection method of designing signal settings and prices for transportation networks. in *Transportation Planning-State of the Art, Proceedings of the 6th Meeting of the EURO Working Group on Transportation*, 9-11 September 1998, Gothenburg, Sweden, M. Patriksson and M. Labbe (eds.), Kluwer Academic Publishers, (1998)

[3] P. Marcotte and D. L. Zhu. Exact and inexact penalty methods for the generalized bilevel programming problem. *Mathematical Programming*, A 74, 141-157 (1996).

[4] J. Outrata and J. Zowe. A numerical approach to optimization problems with variational inequality constraints. *Mathematical Programming*, 68, 105–130 (1995).

[5] M. Patriksson. Nonlinear Programming and Variational Inequality Problems—A Unified Approach, Applied Optimization, Kluwer Academic Publishers, 1999

[6] M.J. Smith. A descent method for solving monotone variational inequalities and monotone complementarity problems. *JOTA* 44, 485–496 (1984).

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38330 Montbonnot-St-Martin (France)
Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399