Wuhan University Journal of Natural Sciences

2010, Vol.15 No.6, 471-475

Article ID 1007-1202(2010)06-0471-05

DOI 10.1007/s11859-010-0686-7

Exact Penalty Method for the Nonlinear Bilevel Programming Problem

□ PAN Qingfei[1], AN Zhonghua[2], QI Hui[3]

1. Department of Physics and Mechanics-Electronic Engineering, Sanming University, Sanming 365004, Fujian, China;

2. Department of Mathematics and Measure Economics, Hubei University of Education, Wuhan 430205, Hubei, China;

3. Department of Mathematics and Computer, Sanming University, Sanming 365004, Fujian, China

Abstract: In this paper, following the method of replacing the lower level problem with its Kuhn-Tucker optimality condition, we transform the nonlinear bilevel programming problem into a normal nonlinear programming problem with the complementary slackness constraint condition. Then, we get the penalized problem of the normal nonlinear programming problem by appending the complementary slackness condition to the upper level objective with a penalty. We prove that this penalty function is exact and the penalized problem and the nonlinear bilevel programming problem have the same global optimal solution set. Finally, we propose an algorithm for the nonlinear bilevel programming problem. The numerical results show that the algorithm is feasible and efficient.

Key words: convex-quadratic programming; nonlinear bilevel programming; Kuhn-Tucker optimality condition; penalty function method; optimal solution

CLC number: O 122.2

Received date: 2010-03-18

Foundation item: Supported by the Key Project on Science and Technology of Hubei Provincial Department of Education (D20103001)

Biography: PAN Qingfei, male, Associate professor, research direction: optimization theory and methods. E-mail:pqf101@yahoo.com.cn

0 Introduction

Among all the approaches for bilevel problems (BP), the penalty function method is one of the most powerful approaches, both as a theoretical tool and as a computational vehicle. In Ref.[1], the linear bilevel programming problem(LBP) is considered and the dual gap function is used as an exact penalty function. Ishizuka *et al*[2] presented a double penalty function method for LBP. Using the uniform parametric error bound as a tool, Ye *et al*[3] established several exact penalty functions for generalized LBP and discussed their relations. Marcotte *et al*[4] discussed exact and inexact penalty functions for generalized LBP. Luo *et al*[5] discussed exact penalty functions for generalized LBP in more general cases and pointed out that it is difficult to get a smooth exact penalty function for LBP. Liu *et al*[6] proposed a new constraint qualification and obtained a smooth exact penalty function for convex LBP. However, most of the researches mentioned above pay more attention to the optimality conditions of LBP but ignore proposing an algorithm for LBP.

In fact, with the increasing extension of the fields in which BP is applied, proposing an algorithm with an attractive computational perspective and numerical stability for the nonlinear BP has become a hot point in the BP field. Note that Colson *et al*[7] have presented a trust-region approach for the general nonlinear BP, and the computational experience shows that the algorithm proposed is rather effective for the nonlinear BP. However, the main deficiency of the algorithm proposed in Ref.[7] is that it is short of the necessary convergence analysis.

In this paper, we consider the nonlinear bilevel problem

$$\text{BP}: \min_x F(x,y) \quad \text{s.t. } G(x) \leqslant 0$$
$$\min_y f(x, y)$$
$$\text{s.t. } g(x, y) \leqslant 0 \tag{1}$$

where F, f: $\mathbf{R}^{m+n} \to \mathbf{R}, G: \mathbf{R}^m \to \mathbf{R}^p, g: \mathbf{R}^{m+n} \to \mathbf{R}^q$ are continuously differentiable functions and $x \in \mathbf{R}^m, y \in \mathbf{R}^n$.

This paper transforms the nonlinear BP to the differentiable nonlinear programming problem equivalently in the sense of a global optimal solution and then propose a simple algorithm for the nonlinear BP. We firstly formulate the penalty problem of (1), and adopt the method of replacing the lower level problem with its Kuhn-Tucker optimality condition. Then we append the complementary slackness condition to the upper level objective with a penalty. We prove that if the lower level problem satisfies some normal constraint qualifications, then BP has the same global optimal solution set with the penalized problem. Although the kind of method for constructing the penalty function in this paper is similar to Refs. [6,8-10], the condition proposed in our thesis is weaker than Ref. [6], and it is easy to verify.

1 Formulation of the Problem

Let $S = \{(x,y) \in \mathbf{R}^{m+n} \mid G(x) \leqslant 0, g(x,y) \leqslant 0\}$ denote the constraint region of problem (1). $S(X) = \{x \in \mathbf{R}^m \mid (x,y) \in S\}$ is the projection of S onto the leader's decision space, $S(x) = \{y \in \mathbf{R}^n \mid g(x,y) \leqslant 0\}$ is the follower's feasible set for every fixed $x \in S(X)$ and $P(x) = \{y \mid y \in \operatorname{agr\,min}[f(x,\bar{y}) \mid \bar{y} \in S(x)]\}$ is the follower's rational reaction set for fixed $x \in S(X)$.

Definition 1 A point (x, y) is said to be feasible to problem (1) if $x \in S(X)$ and $y \in P(x)$. The set of all feasible points is called the inducible region of problem (1) and is denoted by IR, that is to say, $\text{IR} = \{((x,y) \in \mathbf{R}^{m+n}) \mid (x,y) \in S, y \in P(x)\}$. Then, problem (1) can be written as:

$$\min\{F(x,y) \mid (x,y) \in \text{IR}\}$$

Definition 2 A point (x^*, y^*) is said to be an optimal solution of problem (1), if $(x^*, y^*) \in \text{IR}$ and for all $(x,y) \in \text{IR}$, $F(x^*, y^*) \leqslant F(x,y)$.

In order to better define problem (1), throughout this paper, we make the following assumptions.

H_1: The constraint region S is nonempty and compact, $P(x)$ is nonempty and a point-to-point map.

H_2: For fixed $x \in S(X), f(x,\cdot), g(x,\cdot)$ are convex and the lower level problem satisfies the Mangasarian-Fromovitz constraint qualification (MFCQ) at $y \in S(x)$.

Note that the upper level constraint condition $G(x) \leqslant 0$ does not involve the lower level variable y, then the assumption H_1 guarantees that $\text{IR} \neq \varnothing$ [11]. What is more, in assumption H_2, MFCQ can be replaced by any constraint qualification for smooth optimization problems.

If the assumption H_2 is satisfied, we can reduce (1) to the following nonlinear programming problem with the complementary slackness condition

$$\begin{aligned} P: \quad & \min_{x,y,\lambda} F(x,y) \\ & \text{s.t. } G(x) \leqslant 0 \\ & \nabla_y L(x,y,\lambda) = 0 \\ & \lambda^{\mathrm{T}} g(x,y) = 0 \\ & g(x,y) \leqslant 0 \\ & \lambda \geqslant 0 \end{aligned} \tag{2}$$

where $L(x,y,\lambda) = f(x,y) + \lambda^{\mathrm{T}} g(x,y)$ is the Lagrange function and $\lambda \in \mathbf{R}^q$.

Remark 1 If the lower level problem is not a convex parametric optimization problem, then (2) has a larger feasible set including not only global optimal solutions for the lower level problem but also all local optimal solutions and all stationary points. On the other hand, it should also be noted that problem (1) and problem (2) have the same global solution set, and each local optimal solution of problem (1) corresponds to a local optimal solution for problem (2), but the opposite direction is not true in general[12].

For problem (2), due to the existence of the complementary slackness condition, the regularity assumptions which are needed for successfully handling smooth optimization problems are never satisfied. To overcome this difficulty, we append the complementary slackness condition to the upper level objective with a penalty and obtain the following penalized problem.

$$\begin{aligned} P(\mu): \quad & \min_{x,y,\lambda} F(x,y) - \mu\lambda^{\mathrm{T}} g(x,y) \\ & \text{s.t. } G(x) \leqslant 0 \\ & \nabla_y L(x,y,\lambda) = 0 \\ & g(x,y) \leqslant 0 \\ & \lambda \leqslant 0 \end{aligned} \tag{3}$$

where μ is a positive number. In the following section, we will analyze the relationship between the optimal solution of problem (1) and that of the corresponding penalized problem (3), and prove that the penalty function is exact.

2 Main Theoretical Results

For each value of the weight μ, we denote by $(x(\mu), y(\mu), z(\mu))$ a global optimum of $P(\mu)$. The penalty function method is obtained by specifying a sequence of increasing weights $\{\mu_k\}$ and the associated sequence of iterates $\{(x(\mu_k), y(\mu_k), z(\mu_k)\}$. Firstly, we will give the following lemma, which is the essence of the penalty function method, and the convergence result of the algorithm given in Section 3 also relies on it.

Lemma 1 Let the assumptions H_1 and H_2 be satisfied and $\{x_k = x(\mu_k), y_k = y(\mu_k), \lambda_k = \lambda(\mu_k)\}$ be a sequence generated by a penalty algorithm based on the penalty function $P(\mu)$. Let (x^*, y^*) denote an optimal solution to problem (1), set $F^* = F(x^*, y^*)$, then

$$F(x_{k+1}, y_{k+1}) - \mu_{k+1}\lambda_{k+1}^{\mathrm{T}} g(x_{k+1}, y_{k+1}) \geqslant F(x_k, y_k) - \mu_k \lambda_k^{\mathrm{T}} g(x_k, y_k) \quad (4)$$

$$-\lambda_{k+1}^{\mathrm{T}} g(x_{k+1}, y_{k+1}) \leqslant -\lambda_k^{\mathrm{T}} g(x_k, y_k) \quad (5)$$

$$F(x_k, y_k) \leqslant F(x_{k+1}, y_{k+1}) \quad (6)$$

$$F(x_k, y_k) \leqslant F^* \quad (7)$$

Proof Since x_k is optimal for the corresponding penalized program, and μ_k is monotone, we have

$$\begin{aligned} &F(x_k, y_k) - \mu_k \lambda_k^{\mathrm{T}} g(x_k, y_k) \\ &\leqslant F(x_{k+1}, y_{k+1}) - \mu_k \lambda_{k+1}^{\mathrm{T}} g(x_{k+1}, y_{k+1}) \\ &\leqslant F(x_{k+1}, y_{k+1}) - \mu_{k+1} \lambda_{k+1}^{\mathrm{T}} g(x_{k+1}, y_{k+1}) \end{aligned} \quad (8)$$

Then the inequality (4) is exact. Similarly, we have

$$\begin{aligned} &F(x_{k+1}, y_{k+1}) - \mu_{k+1}\lambda_{k+1}^{\mathrm{T}} g(x_{k+1}, y_{k+1}) \\ &\leqslant F(x_k, y_k) - \mu_{k+1}\lambda_k^{\mathrm{T}} g(x_k, y_k) \end{aligned}$$

Combining the above two inequalities, we yield

$$(\mu_{k+1} - \mu_k)\lambda_{k+1}^{\mathrm{T}} g(x_{k+1}, y_{k+1}) \geqslant (\mu_{k+1} - \mu_k)\lambda_k^{\mathrm{T}} g(x_k, y_k)$$

Then we have

$$-\lambda_{k+1}^{\mathrm{T}} g(x_{k+1}, y_{k+1}) \leqslant -\lambda_k^{\mathrm{T}} g(x_k, y_k)$$

Combining the above inequalities with (8), we yield (6). Finally, we observe that

$$\begin{aligned} &F^* = F(x^*, y^*) - \mu_k \lambda_k^{\mathrm{T}} g(x^*, y^*) \\ &\geqslant F(x_k, y_k) - \mu_k \lambda_k^{\mathrm{T}} g(x_k, y_k) \geqslant F(x_k, y_k) \end{aligned}$$

Then the proof is completed.

We now show that a finite value of μ would yield an exact solution to the overall problem (3), where the penalty element $\lambda^{\mathrm{T}} g(x, y)$ becomes zero; moreover, problems P and $P(\mu)$ have the same nonempty global solution set.

Let $Z = \{(x, y, \lambda) \mid G(x) \leqslant 0, \nabla_y L(x, y, \lambda) = 0, g(x, y) \leqslant 0, \lambda \geqslant 0\}$, $\aleph(x, y) = \{\lambda \mid \nabla_y L(x, y, \lambda) = 0, \lambda \geqslant 0\}$ and λ_i, $(i = 1, \cdots, k)$ denote the extreme points of $\aleph(x, y)$, and let $V[P]$ and $V[P(\mu)]$ denote the optimal value of problems P and, $P(\mu)$ respectively. Then, inspired from the method presented in Refs. [8, 13], we have the following theorem.

Theorem 1 Let the assumptions H_1 and H_2 be satisfied, then there exists $\mu^* \geqslant \mu^0$ (some finite value) such that:

(i) There exists (x^*, y^*, λ^*), which is feasible to problem $P(\mu)$, satisfying $(\lambda^*)^{\mathrm{T}} g(x^*, y^*) = 0$ and $V[P] = V[P(\mu)] = F(x^*, y^*)$.

(ii) Moreover, P and $P(\mu)$ have the same nonempty global solution set for all $\mu \geqslant \mu^*$.

Proof (i) Following the assumptions H_1 and H_2, it is obvious that problem P is feasible and $P(\mu_0)$ has a solution for some $\mu_0 \geqslant 0$. As $P(\mu)$ is less restricted than P and $-\lambda^{\mathrm{T}} g(x, y) \geqslant 0$, for all feasible points of $P(\mu)$, $P(\mu)$ has a solution for all $\mu \geqslant \mu_0$.

Then we have

$$V[P(\mu)] = \min\{F(x, y) - \mu\lambda^{\mathrm{T}} g(x, y) \mid (x, y, \lambda) \in Z\}, \quad \mu \geqslant \mu_0 \quad (9)$$

Let us split Z into the two subsets, $C_0 = \{(x, y, \lambda) \in Z \mid \lambda^{\mathrm{T}} g(x, y) = 0\}$ and $C_1 = Z / C_0$.

If $C_1 \neq \varnothing$, we will firstly prove that the value of $\max\{(F(x^*, y^*) - F(x, y)) / -\lambda^{\mathrm{T}} g(x, y) \mid (x, y, \lambda) \in C_1\}$ is finite.

For $\lambda \in \aleph(x, y)$, we have $\lambda = \sum_{i=1}^{k} \alpha_i \lambda_i$ where $\alpha_i \geqslant 0$ and $\sum_{i=1}^{k} \alpha_i = 1$.

Following the assumption H_1, it is obvious that $D(\lambda) = \{(x, y) \mid (x, y, \lambda) \in C_1\}$ is nonempty and compact. Then, the value of $\min_i \{-\lambda_i^{\mathrm{T}} g(x, y) \mid (x, y) \in D(\lambda)\}$ is finite. Let $\min_i \{-\lambda_i^{\mathrm{T}} g(x, y) \mid (x, y) \in D(\lambda)\} = m$ (positive constant). Then

$$\begin{aligned} &\max\left\{\frac{(F(x^*, y^*) - F(x, y))}{-\lambda^{\mathrm{T}} g(x, y)} \middle| (x, y, \lambda) \in C_1\right\} \\ &\leqslant \frac{\max\{F(x^*, y^*) - F(x, y) \mid (x, y) \in D(\lambda)\}}{\min\{-\lambda^{\mathrm{T}} g(x, y) \mid (x, y, \lambda) \in C_1\}} \\ &= \frac{\max\{F(x^*, y^*) - F(x, y) \mid (x, y) \in D(\lambda)\}}{\min\left\{-\sum_{i=1}^{k} \alpha_i \lambda_i^{\mathrm{T}} g(x, y) \middle| (x, y, \lambda) \in C_1\right\}} \\ &\leqslant \frac{\max\{F(x^*, y^*) - F(x, y) \mid (x, y) \in D(\lambda)\}}{\sum_{i=1}^{k} \alpha_i m} \end{aligned}$$

$$=\frac{\max\{F(x^*,y^*)-F(x,y)|(x,y)\in D(\lambda)\}}{m}$$

Then, it means that the value of $\max\{(F(x^*,y^*)-F(x,y))/-\lambda^{\mathrm{T}}g(x,y)|(x,y,\lambda)\in C_1\}$ is finite.

Then, we can take

$$\mu^*=\max\{\mu_0,\max\{(F(x^*,y^*)-F(x,y))/-\lambda^{\mathrm{T}}g(x,y)|(x,y,\lambda)\in C_1\}\} \quad (10)$$

Thus for all $\mu\geqslant\mu^*$, we have

$$L(\mu)=\min\{F(x,y)-\mu\lambda^{\mathrm{T}}g(x,y)|(x,y,\lambda)\in C_1\}\geqslant F(x^*,y^*) \quad (11)$$

Following (9)-(11) and taking into account that $\mu^*\geqslant\mu_0$, we have

$$V[P(\mu)]=\min\{F(x^*,y^*),L(\mu)\}=F(x^*,y^*),\ \mu\geqslant\mu^* \quad (12)$$

By $C_1\neq\varnothing$, it is obvious that

$$V[P(\mu)]=F(x^*,y^*)\ \ \mu\geqslant\mu^*=\mu_0$$

So, it results in $(\lambda^*)^{\mathrm{T}}g(x^*,y^*)=0$ and

$$V[P]=V[P(\mu)]=F(x^*,y^*)\ \ \mu\geqslant\mu^* \quad (13)$$

(ii) Let $\bar{\mu}\geqslant\mu^*$. Let us denote by $O(P)$ and $O(P(\bar{\mu}))$ the optimal sets of problems P and $P(\bar{\mu})$. By (12), we have that $O(P)\neq\varnothing$ and $O(P(\bar{\mu}))\neq\varnothing$ with $V[P]=V[P(\bar{\mu})]$. Since any feasible solution of P is feasible to $P(\bar{\mu})$, then $O(P)\subseteq O(P(\bar{\mu}))$.

Conversely, let $(x,y,\lambda)\in O(P(\bar{\mu}))$. Following (12) together with $\bar{\mu}\geqslant\mu^*$, we have

$$V[P]=V[P(\mu)]=F(x,y)-\bar{\mu}\lambda^{\mathrm{T}}g(x,y)$$
$$\geqslant F(x,y)-\mu^*\lambda^{\mathrm{T}}g(x,y)=V[P(\mu^*)]=V[P]$$

Thus, equality holds among the relations. It yields that $(\bar{\mu}-\mu^*)\lambda^{\mathrm{T}}g(x,y)=0$. Then $\lambda^{\mathrm{T}}g(x,y)=0$. Therefore, (x,y,λ) is feasible to P and $V[P]=F(x,y)$. That is, $(x,y,\lambda)\in O(P)$. Hence, $O(P(\bar{\mu}))\subseteq O(P)$. The proof is completed.

3 The Algorithm and Convergence Analysis

Algorithm

Step 1 Choose $\mu_0>0$ (μ_0 large enough) and $l>0$, $k=0$;

Step 2 Following the Kuhn-Tucker optimality of the lower level problem, transform (1) and obtain the corresponding penalized problem $P(\mu_k)$;

Step 3 Solve the problem $P(\mu_k)$ for its solution (x_k,y_k,λ_k);

Step 4 If $\lambda_k^{\mathrm{T}}g(x_k,y_k)=0$, then stop, the optimal solution of (1) is (x_k,y_k); else, set $\mu_k:=\mu_k+l,k:=k+1$ and return to Step 3.

Theorem 2 Let the assumptions H_1 and H_2 be satisfied, then the set $\{(x_k,y_k)\}$ coming from the algorithm converges to the optimal solution of the BLP problem (1).

Proof By Theorem 1, we have that the penalty function is exact. Then, we only need to prove that the last point in the set $\{(x_k,y_k)\}$ solves problem (1). Let the algorithm terminate at μ_k and the solution be (x_k,y_k). Then following the algorithm, we have $\lambda_k^{\mathrm{T}}g(x_k,y_k)=0$.

If the term (x_k,y_k) does not solve problem (1), then following Theorem 1, there must exist some $\mu^*>\mu_k$ such that the term (x^*,y^*) solves problem (1). Then we have $(\lambda^*)^{\mathrm{T}}g(x^*,y^*)=0$ and

$$F(x_k,y_k)>F(x^*,y^*) \quad (14)$$

However, as $\mu^*>\mu_k$, following the Lemma it should be

$$F(x_k,y_k)-\lambda_k^{\mathrm{T}}g(x_k,y_k)=F(x_k,y_k)$$
$$\leqslant F(x^*,y^*)-(\lambda^*)^{\mathrm{T}}g(x^*,y^*)=F(x^*,y^*) \quad (15)$$

The term (14) contradicts with (15), and this contradiction proves Theorem 2.

4 Numerical Experiments

Now, we use four examples to validate our algorithm. We make a program with MATLAB language and use a personal computer (CPU: Intel Pentium 1.7 GHz, RAM: 256 MB) to execute the program. The values of the parameters are $\mu_0=50,l=10$. The numerical results are shown in Table 1.

Test problems of the convex-quadratic type are as follows (variable $x=(x_1x,x_2,\cdots,x_m)\in\mathbf{R}^m$, $y=(y_1,y_2,\cdots,y_n)\in\mathbf{R}^n$).

Example 1: Bard98Ex1[14]

$$\min_{x\geqslant0}(x-5)^2+(2y+1)^2$$
$$\min_{y\geqslant0}(y-1)^2-1.5xy$$
$$\text{s.t.}\ \ -3x+y+3\leqslant0$$
$$x-0.5y-4\leqslant0$$
$$x+y-7\leqslant0$$

Example 2: Bard98Ex2[14]

$$\min_x 0.5(x_1-0.8)^2+0.5(x_2-0.5)^2+0.5(y-1)$$
$$\text{s.t.}\ \ 0\leqslant x_1,x_2\leqslant1$$
$$\min_y 0.5y^2+y-x_1y+2x_2y$$
$$\text{s.t.}\ \ 0\leqslant y\leqslant1$$

Example 3: Muu03Ex3[15]

$$\min_x x^2-4x+y_1^2+y_2^2$$

s.t. $0 \leqslant x \leqslant 2$

$\min_{y} y_1^2 + 0.5y_2^2 + y_1y_2 + (1-3x)y_1 + (1+x)y_2$

s.t. $2y_1 + y_2 - 2x \leqslant 1$, $y \geqslant 0$

Example 4: Muu03Ex4 [15]

$\min_{x\geqslant 0} y_1^2 + y_3^2 - y_1y_3 - 4y_2 - 7x_1 + 4x_2$

s.t. $x_1 + x_2 \leqslant 1$

$\min_{y\geqslant 0} y_1^2 + 0.5y_2^2 + 0.5y_3^2 + y_1y_2 + (1-3x_1)y_1 + (1+x_2)y_2$

s.t. $2y_1 + y_2 - y_3 + x_1 - 2x_2 + 2 \leqslant 0$

Table 1 Results of test problems with MATLAB

Problem name	m	n	Iteration	CPU time/s	Optimal solution (x, y)
Bard98Ex1	1	1	1	0.45	(1,0)
Bard98Ex2	2	1	1	0.36	(0.846, 0.769, 0)
Muu03Ex3	1	2	1	1.26	(1,0,1)
Muu03Ex4	2	3	1	1.78	(0.611, 0.389, 0, 0, 1.83)

In fact, for the convex-quadratic BLP problem, all test problems in Ref.[15] are solved using the algorithm proposed in this paper. In Table 1, we can find that all problems are solved in one iteration. This shows that our algorithm is simple and effective. The reason is that we adopt a large enough penalty parameter μ_0.

5 Conclusion

We explore the equivalent relationship between the nonlinear BLP problem and the corresponding differential nonlinear programming problem in the sense of a global optimal solution. The numerical results reflect the good behavior of the algorithm proposed in this paper. Firstly, the algorithm is effective especially for convex-quadratic BP because of the fact that the corresponding nonlinear programming problem $P(\mu)$ is still a convex programming problem. Secondly, the number of test problems treated in this paper is small. This is due to the fact that very few instances of large-scale nonlinear BP are available in the literature. Moreover, we give the strict convergence theorem of the algorithm, and then it guarantees that for the large-scale problem we can still get its optimal solution. Finally, exploring the relationship between the local optimal solution of the nonlinear BP and that of the penalized problem $P(\mu)$ will be our future work.

References

[1] Anandalingam G, White D J. A solution for the linear static Stackelberg problem using penalty function[J]. *IEEE Transactions Automatic Control*, 1990, **35**: 1170-1173.

[2] Ishizuka Y, Aiyoshi E. Double penalty method for bilevel optimization problems[J]. *Annals of Operations Research*, 1992, **34**: 73-88.

[3] Ye J, Zhu D L, Zhu Q. Generalized bilevel programming problems[J]. *SIAM Journal on Optimization*, 1997, **33**: 481- 507.

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

[5] Luo Z Q, Pang J S, Ralph D, *et al*. Exact penalization and stationarity conditions of mathematics programs with equilibrium constraints[J]. *Mathematical Programming*, 1996, **75**: 19-76.

[6] Liu G S, Han J Y, Zhang J Z. Exact penalty function for convex bilevel programming problems[J]. *Journal of Optimization Theory and Applications*, 2001, **110**(3): 621-643.

[7] Colson B, Marcotte P, Savard G. A trust-region method for nonlinear bilevel programming: algorithm and computational experience[J]. *Computational Optimization and Applications*, 2005, **30**(3): 211-227.

[8] Campelo M, Scheimberg S. A study of local solutions in linear bilevel programming[J]. *Journal of Optimization Theory and Application*, 2005, **125**(1): 63-84.

[9] Lv Y, Hu T, Wang G, *et al.* A penalty function method based on Kuhn-Tucker condition for solving linear bilevel programming[J]. *Applied Mathematics and Computation*, 2007, **188**: 808-813.

[10] Lv Y, Hu T, Wan Z. A penalty function method for solving inverse optimal value problem[J]. *Journal of Computational and Applied Mathematics*, 2008, **220**(1-2): 175-180.

[11] Shi C, Zhang G, Lu J. On the definition of linear bilevel programming solution[J]. *Applied Mathematics and Computation*, 2005, **160**: 169-176.

[12] Dempe S. *Foundation of Bilevel Programming*[M]. London: Kluwer Academic Publishers, 2002.

[13] Campelo M, Scheimberg S. Theoretical and computational results for a linear bilevel problem[J]. *Multilevel Optimization: Algorithm and Applications*, 2001, **54**: 269-281.

[14] Bard J. *Practical Bilevel Optimization: Algorithm and Applications*[M]. Dordrecht : Kluwer Academic Publishers, 1998.

[15] Muu L D, Quy N V. A global optimization method for solving convex quadratic bilevel programming problems[J]. *Journal of Global Optimization*, 2003, **26**: 199-219.

□