A TRUST REGION ALGORITHM FOR NONLINEAR BILEVEL PROGRAMMING

P. Marcotte *
G. Savard †
D. L. Zhu ‡

April 4, 2002

Abstract We propose to solve generalized bilevel programs by a trust region approach where the "model" takes the form of a bilevel program involving a linear program at the upper level and a linear variational inequality at the lower level. By coupling the concepts of trust region and linesearch in a novel way, we obtain an implementable algorithm that converges to a strong stationary point of the original bilevel program.

Keywords: Bilevel programming. Trust region. Linesearch.

1 A trust region approach to bilevel programming

In this paper we consider bilevel programs (or MPECs, i.e., mathematical programs with equilibrium constraints, see Luo, Pang and Ralph [5]) of the form

$$\begin{aligned} \min_{x\in X, y\in Y(x)} \quad & f(x,y) \\ \text{s.t.} \quad & \langle F(x,y), y-y'\rangle \le 0 \quad \forall y' \in Y(x) \end{aligned}$$

where the mapping F is strongly monotone with respect to the lower level variable y and where the sets X and $Y(x) = \{y : Ax + By \ge b\}$ are polyhedral. We propose for its solution a trust region approach where the model is itself a bilevel program of a combinatorial nature that can be solved for a global optimum.

This approach, which mixes continuous and discrete optimization, trust regions and linesearches, is in stark contrast with more traditional descent methods whose convergence properties are frequently weak and difficult to analyze (see the works of Outrata, Kočvara and Zowe [7] or, more recently, the sequential programming scheme developed by Fukushima and Tseng [3].) Although the underlying motivation is quite different, it is related to a trust

*DIRO and CRT, Université de Montréal, CP 6128, succursale Centre-ville, Montréal, QC, Canada H3C 3J7 (correspondance author)
†GERAD and Département de mathématiques et génie industriel, École Polytechnique, C.P. 6079, succursale Centre-ville, Montréal, QC, Canada H3C 3A7
‡CRT, Université de Montréal, CP 6128, succursale Centre-ville, Montréal, QC, Canada H3C 3J7 and Fudan University, Shanghai, China

region method applied to an exact penalty reformulation of the original bilevel program, such as proposed by Scholtes and Stöhr [10].

The core of the algorithm consists in solving, at each iteration, the affine approximation of the original bilevel model obtained by replacing f and F by their respective first-order developments $\bar{f}$ and $\bar{F}$ around the current iterate $(\bar{x}, \bar{y})$, and forcing the upper level decision vector x to lie within some distance ϵ (the radius of the trust region) of the current iterate, i.e.,

$$\begin{array}{ll} \min\limits_{x \in X, y \in Y(x)} & \bar{f}(x, y) \\ \text{s.t.} & \langle \bar{F}(x, y), y - y' \rangle \leq 0 \quad \forall y' \in Y(x) \\ & \|x - \bar{x}\| \leq \epsilon. \end{array}$$

Starting from $(\bar{x}, \bar{y})$, the algorithm sets the next iterate to a solution $(\bar{x}_\epsilon, \bar{y}_\epsilon)$ of the above linearized program, provided that the trust region model is deemed a suitable approximation to the original bilevel program, i.e., that the ratio

$$\frac{f(\bar{x}, y(\bar{x})) - f(\bar{x}_\epsilon, y(\bar{x}_\epsilon))}{\bar{f}(\bar{x}, \bar{y}(\bar{x})) - \bar{f}(\bar{x}_\epsilon, \bar{y}(\bar{x}_\epsilon))}$$

where $y(x)$ and $\bar{y}(x)$ represent the well-defined optimal reaction functions, is bounded away from zero. If the ratio is small, the iterate is left unchanged and the radius of the trust region is decreased; otherwise the point $\bar{x}_\epsilon$ replaces the iterate $\bar{x}$ as the new iterate. This yields the following classical trust region algorithm which is, by construction, a descent algorithm.

Algorithm BlTrust

Step 0 [Initialization]

- Select positive parameters ϵ_0 and an initial iterate $x^0 \in X$.
- Set the iteration index k to 0.

Step 1

$\bar{x}_{\epsilon_k} \leftarrow$ solution of the linearized problem within the trust region.

Step 2

Set $r_k = \dfrac{f(x^k, y(x^k)) - f(\bar{x}_{\epsilon_k}, y(\bar{x}_{\epsilon_k}))}{\bar{f}(x^k, \bar{y}(x^k)) - \bar{f}(\bar{x}_{\epsilon_k}, \bar{y}(\bar{x}_{\epsilon_k}))}$.

if $2/3 \leq r_k$ **then** $x^{k+1} \leftarrow \bar{x}_{\epsilon_k}$ and $\epsilon_{k+1} \leftarrow 2\epsilon_k$
if $1/3 \leq r_k < 2/3$ **then** $x^{k+1} \leftarrow \bar{x}_{\epsilon_k}$ and $\epsilon_{k+1} \leftarrow \epsilon_k$
if $r_k < 1/3$ **then** $x^{k+1} \leftarrow x^k$ and $\epsilon_{k+1} \leftarrow \epsilon_k/2$

Step 3

- $k \leftarrow k + 1$
- Return to Step 1.

At **Step 1** of the algorithm, the lower level linear variational inequality can be replaced by its Karush-Kuhn-Tucker system, to yield the equivalent primal-dual single-level program

$$
\begin{array}{ll}
\min\limits_{x} & \bar{f}(x,y) \\
 & \\
\text{s.t.} & x \in X \\
 & \|x-\bar{x}\| \leq \epsilon \\
 & \\
 & Ax+By \geq b \\
 & \pi B = \bar{F}(x,y) \\
 & \pi \geq 0 \\
 & \pi(Ax+By-b)=0.
\end{array}
$$

Upon introduction of a row vector of binary variables z, the sole nonlinear constraint of the above program can be written as

$$
\begin{array}{rcl}
\pi & \leq & Mz \\
Ax+By-b & \leq & M(e-z^T)
\end{array}
$$

where the components of the vector e are all equal to one and where M is a suitably large (finite) constant which also appears in exact penalty reformulations of the bilevel program (see Marcotte and Zhu [6] or Scholtes and Stöhr [10]). Whenever the norm $\|\cdot\|$ is polyhedral, this mixed-integer program can be solved exactly for a global optimum by implicit enumeration (Branch-and-Bound) techniques.

Algorithm BlTrust is closely related to the trust region method applied in [10] to an exact penalization of a mathematical program with equilibrium constraints (MPEC) and where the "model" is nonsmooth. The difference between the approaches is twofold:

- Scholtes and Stöhr apply the trust region constraint to both the upper and lower level variables. Actually, in an MPEC formulation of the bilevel program, their trust region applies as well to the lower level dual vector;

- no solution algorithm for the nonsmooth model is specified whereas, in our case, the model can be solved exactly, by an algorithm similar to that of Hansen, Jaumard and Savard [4]. If both upper and lower level functions are mildly nonlinear, we expect that the trust region model will be in close agreement with the original program, and that a nearly globally optimal solution will be uncovered in the early stages of the procedure.

Under standard assumptions on the functions f, F and the constraint sets X and $Y(x)$, it is usually only possible to prove that the limit points of the sequence $\{x^k\}$ satisfy first-order optimality conditions in the Clarke sense (see Clarke [1]). The blanket assumption required to prove the stronger result that limit points are B-stationary, i.e., that there do not exist feasible descent directions, is that the sequence of radii ϵ_k stays bounded away from zero. Rather than making such a claim, we will modify the last statement at **Step 2** of the basic algorithm to incorporate a linesearch strategy that will guarantee that limit points are B-stationary, thus eliminating the requirement about trust region radii.

$$\textbf{if} \quad r_k < 1/3 \quad \textbf{then} \quad \begin{array}{l} \nu_k \in \{\arg \min_{\nu \in \{2^j \epsilon_k : 1 \leq j \leq \lceil -\lg \epsilon_k \rceil\}} f(\bar{x}_\nu, y(\bar{x}_\nu))\} \\ x^{k+1} \in \arg\min\{f(\bar{x}_{\nu_k}, y(\bar{x}_{\nu_k}), f(x^k, y(x^k))\} \\ \epsilon_{k+1} \leftarrow \epsilon_k/2 \end{array}$$

The rationale underlying the linesearch rule is that, in the vicinity of a nonstationary point, there must exist a trust region radius ϵ such that $\bar{x}_\epsilon$ improves on the current iterate x^k, even though the ratio r_k indicates that the linear model is a poor approximation of the original model; by doubling the trust region radius a finite number of times, one is able to bypass such nonstationary points. While this scheme may appear computationally costly at a first glance, it is not unreasonable, considering that (i) "tractable" bilevel programs must involve lower level problems that can be efficiently and repeatedly solved (ii) the probability of having to perform the linesearch is low, as confirmed by our numerical experiments.

We now present the key assumptions that will be used throughout the paper.

Assumption 1 *The set X is compact.*

Assumption 2 *The function f and its gradient ∇f are Lipschitz continuous on $Z = \{(x, y) : x \in X, y \in Y(x)\}$, with respective Lipschitz constants l and l'.*

Assumption 3 *The mapping F and its Jacobian are Lipschitz continuous on Z, with respective Lipschitz constants L and L'.*

Assumption 4 *F is uniformly strongly monotone in y, with modulus β, on Z.*

Note that the above Lipschitz and monotonicity properties are shared by the first-order approximations of f and F, respectively.

The remainder of the paper will be divided into four parts: the local behaviour of the reaction functions $y(x)$ and $\bar{y}(x)$, a series of approximation lemmas, the convergence analysis and a conclusion.

2 Local behaviour of $y(x)$ and $\bar{y}(x)$

Theorem 1 *The functions $y(x)$ and $\bar{y}(x)$ are Lipschitz continuous on their common domain of definition $D = \{x \in X : Y(x) \neq \emptyset\}$.*

Proof. Under local Lipschitz and strong monotonicity assumptions on the mapping F, Dafermos [2] has shown that the reaction function $y(x)$ is Lipschitz continuous, provided that the projection operator $p_z(x) = \mathrm{proj}_{Y(x)}(z)$ is also Lipschitz continuous, for fixed z. By definition, we have that

$$p_z(x) = \arg \min_{y \in Y(x)} \frac{1}{2} \|y - z\|^2$$

or equivalently, $p_z(x)$ satisfies the generalized equation

$$0 \in p_z(x) - z + \partial\delta(Y(x)),$$

where δ denotes the indicator function of the polyhedral set $Y(x)$. It follows from Robinson [9] that the polyhedral multifunction $p_z(x)$ is upper Lipschitzian. Since the set X is compact and $p_z(x)$ is uniquely defined, we conclude that $p_z(x)$ is Lipschitz continuous on its domain, and this brings the desired result. An identical proof applies to the reaction function $\bar{y}(x)$. □

Theorem 2 *The functions $y(x)$ and $\bar{y}(x)$ are directionally differentiable.*

Proof. Let $c = F(x, y(x))$. Consider the linear program

$$\begin{array}{rl} \min\limits_{y} & cy \\ \text{subject to} & Ax + By \geq b \end{array}$$

and its dual

$$\begin{array}{rl} \max\limits_{\pi \geq 0} & \pi(b - Ax) \\ \text{subject to} & \pi B = c. \end{array}$$

The vector $y(x)$ is the optimal solution of the primal program while the optimal set of dual vectors $\pi(x)$ is upper Lipschitzian as a function of the couple (c, x). Since $c = F(x, y(x))$ is Lipschitz continuous, it follows that the multifunction $\pi(x)$ is upper Lipschitzian. This guarantees, for any feasible direction d and any real sequence $\{t_i\}_{i \in I}$ converging to zero, the existence of vectors $\tilde{\pi}(x) \in \pi(x)$, $\tilde{\pi}(x + t_i d) \in \pi(x + t_i d)$ and of a positive number M such that

$$\|\tilde{\pi}(x + t_i d) - \tilde{\pi}(x)\| \leq M t_i \qquad \forall i.$$

Let

$$\pi' = \lim_{i \in I'} \frac{\tilde{\pi}(x + t_i d) - \tilde{\pi}(x)}{t_i}$$

for some subset of indices I' and let us consider, as in Qiu and Magnanti [8], the index sets

$$\begin{array}{rcl} K_1 & = & \{k | \tilde{\pi}(\bar{x}) > 0, (By(\bar{x}) + A\bar{x} - b)_k = 0\}, \\ K_2 & = & \{k | \tilde{\pi}(\bar{x}) = 0, (By(\bar{x}) + A\bar{x} - b)_k = 0\}, \\ K_3 & = & \{k | \tilde{\pi}(\bar{x}) = 0, (By(\bar{x}) + A\bar{x} - b)_k > 0\}. \end{array}$$

Let y' be a limit point of the sequence $\{[y(x + t_i d) - y(x)]/t_i\}_{i \in I}$. According to results of Qiu and Magnanti we obtain that the limit vectors y' and π' satisfy the linear complementarity system

$$\begin{array}{rcll} \pi' B - \nabla_y F(x, y) y' - \nabla_x F(x, y) d & = & 0 & \\ (By')_k & = & 0 & \text{if } k \in K_1 \\ (By')_k & = & 0 & \text{if } k \in K_2 \text{ and } (\pi')_k > 0 \\ (By')_k & \geq & 0 & \text{if } k \in K_2 \text{ and } (\pi')_k = 0 \\ (\pi')_k & \geq & 0 & \text{if } k \in K_2 \\ (\pi')_k & = & 0 & \text{if } k \in K_3. \end{array}$$

The partial solution y' of the above system, which coincides with the directional derivative $y'(x;d)$, is also the unique solution of the strongly monotone variational inequality

$$\langle \nabla_x F(x,y)d + \nabla_y F(x,y)y, y - z\rangle \leq 0 \qquad \forall z \in P,$$

where $P = \{y : (By)_k = 0, k \in K_1, (By)_k \geq 0, k \in K_2\}$. $\square$

The above results justify the local approximation of the function $f(x, y(x))$ by the function $\bar{f}(x, \bar{y}(x))$, since not only $\bar{y}(\bar{x}) = y(\bar{x})$ and $\bar{y}'(\bar{x}; d) = y'(\bar{x}; d)$ but also both lower level variational inequalities, original and approximated, have identical KKT multipliers and directional derivatives.

3 Approximation results

In this section we prove that, applied to an iterate sufficiently close to a point which is not a strong stationary point, i.e., a point where at least one feasible descent direction exists, algorithm BlTrust will generate a new iterate with an objective value strictly less than that of the current iterate. The proof of this result is based on three approximation lemmas, which make use of the notation described in Table 1 below.

ϵ:	radius of the trust region
$y(x)$:	solution of the lower level variational inequality
$(\bar{x}, \bar{y})$:	current feasible iterate: $\bar{y} = y(\bar{x})$
$\bar{y}(x)$:	solution of the linearized lower level variational inequality
$\bar{f}(x, y)$:	linear approximation of f at $(\bar{x}, \bar{y})$
$\bar{F}(x, y)$:	linear approximation of F at $(\bar{x}, \bar{y})$
$(\bar{x}_\epsilon, \bar{y}_\epsilon = \bar{y}(\bar{x}_\epsilon))$:	a global minimizer of the linearized bilevel program (model) within the trust region.
$(x_\epsilon, y(x_\epsilon))$:	a global minimizer of the bilevel program within the trust region
α:	joint Lipschitz constant of $y(x)$ and $\bar{y}(x)$
β:	uniform modulus of strong monotonicity of F with respect to its second argument y
σ:	$\sigma = \frac{1}{2}L'(1+\alpha^2)/\beta$
μ:	$\mu = l\sigma + \frac{1}{2}l'(1+\alpha^2)$
τ:	$\tau = \mu + \frac{1}{2}l'(1+\alpha^2) + l\sigma$
$\tilde{f}$:	$\tilde{f}(x) = f(x, y(x))$

Table 1: Notation

Lemma 1 $\|y(x_\epsilon) - \bar{y}(x_\epsilon)\| \leq \sigma\epsilon^2$ *and* $\|y(\bar{x}_\epsilon) - \bar{y}_\epsilon\| \leq \sigma\epsilon^2$.

Proof. By definition of $y(x_\epsilon)$ and $\bar{y}(x_\epsilon)$ one has

$$\langle F(x_\epsilon, y(x_\epsilon)), y(x_\epsilon) - \bar{y}(x_\epsilon)\rangle \leq 0$$

and

$$\langle \bar{F}(x_\epsilon, \bar{y}(x_\epsilon)), \bar{y}(x_\epsilon) - y(x_\epsilon)\rangle \leq 0.$$

Subtracting those inequalities, one gets

$$\langle F(x_\epsilon, y(x_\epsilon)) - \bar{F}(x_\epsilon, \bar{y}(x_\epsilon)), y(x_\epsilon) - \bar{y}(x_\epsilon)\rangle \leq 0.$$

Adding and subtracting the term $F(x_\epsilon, \bar{y}(x_\epsilon))$ one obtains

$$\langle F(x_\epsilon, y(x_\epsilon)) - F(x_\epsilon, \bar{y}(x_\epsilon)) + F(x_\epsilon, \bar{y}(x_\epsilon)) - \bar{F}(x_\epsilon, \bar{y}(x_\epsilon)), y(x_\epsilon) - \bar{y}(x_\epsilon)\rangle \leq 0$$

and, using the strong monotonicity and Lipschitz continuity of F':

$$\begin{aligned}
\beta\|y(x_\epsilon) - \bar{y}(x_\epsilon)\| &\leq \|F(x_\epsilon, \bar{y}(x_\epsilon)) - \bar{F}(x_\epsilon, \bar{y}(x_\epsilon))\| \\
&\leq \frac{1}{2}L'[\|x_\epsilon - \bar{x}\|^2 + \|\bar{y}(x_\epsilon) - \bar{y}\|^2] \qquad \text{by Theorem 1} \\
&\leq \frac{1}{2}L'[\epsilon^2 + (\alpha\epsilon)^2] \\
&= \beta\sigma\epsilon^2.
\end{aligned}$$

The result follows upon division of each side of the inequality by β. The proof of the second inequality is similar. □

Lemma 2 *The difference between the global minimum within the trust region and the true minimum is of the order of ϵ^2. More precisely:*

$$\bar{f}(\bar{x}_\epsilon, \bar{y}_\epsilon) - f(x_\epsilon, y(x_\epsilon)) \leq \mu\epsilon^2.$$

Proof. Since $(\bar{x}_\epsilon, \bar{y}_\epsilon)$ is optimal for the model, we have

$$\begin{aligned}
\bar{f}(\bar{x}_\epsilon, \bar{y}_\epsilon) &\leq \bar{f}(x_\epsilon, \bar{y}(x_\epsilon)) \\
&\leq f(x_\epsilon, \bar{y}(x_\epsilon)) + \frac{1}{2}l'[\|x_\epsilon - \bar{x}\|^2 + \|\bar{y}(x_\epsilon) - \bar{y}\|^2] \\
&\leq f(x_\epsilon, \bar{y}(x_\epsilon)) + \frac{1}{2}l'\epsilon^2(1 + \alpha^2) \qquad \text{by Theorem 1} \\
&\leq f(x_\epsilon, y(x_\epsilon)) + l\|\bar{y}(x_\epsilon) - y(x_\epsilon)\| + \frac{1}{2}l'\epsilon^2(1 + \alpha^2) \\
&\leq f(x_\epsilon, y(x_\epsilon)) + l\sigma\epsilon^2 + \frac{1}{2}l'\epsilon^2\left(1 + \alpha^2\right) \qquad \text{by Lemma 1} \\
&= f(x_\epsilon, y(x_\epsilon)) + \mu\epsilon^2.
\end{aligned}$$

□

Lemma 3 *The difference between the value predicted by the model and the actual value of the cost function is of the order of ϵ^2. More precisely:*

$$f(\bar{x}_\epsilon, y(\bar{x}_\epsilon)) - f(\bar{x}_\epsilon, \bar{y}_\epsilon) \leq l\sigma\epsilon^2.$$

Proof. $f(\bar{x}_\epsilon, y(\bar{x}_\epsilon)) \leq f(\bar{x}_\epsilon, \bar{y}_\epsilon) + l\|y(\bar{x}_\epsilon) - \bar{y}_\epsilon\| \leq f(\bar{x}_\epsilon, \bar{y}_\epsilon) + l\sigma\epsilon^2$ by Lemma 1. □

Lemma 4 *The difference between the best value of the cost function within the trust region and the actual value given by the model is of the order of ϵ^2. More precisely:*

$$|\tilde{f}(\bar{x}_\epsilon) - \tilde{f}(x_\epsilon)| \leq \tau\epsilon^2.$$

Proof.

$$\begin{aligned}
\tilde{f}(\bar{x}_\epsilon) &= f(\bar{x}_\epsilon, y(\bar{x}_\epsilon)) \\
&\leq f(\bar{x}_\epsilon, \bar{y}(\bar{x}_\epsilon)) + l\sigma\epsilon^2 && \text{by Lemma 3} \\
&\leq \bar{f}(\bar{x}_\epsilon, \bar{y}(\bar{x}_\epsilon)) + \frac{1}{2}l'[\|\bar{x}_\epsilon - \bar{x}\|^2 + \|\bar{y}_\epsilon - \bar{y}\|^2] + l\sigma\epsilon^2 \\
&\leq \bar{f}(\bar{x}_\epsilon, \bar{y}(\bar{x}_\epsilon)) + \frac{1}{2}l'[(1+\alpha^2)\|\bar{x}_\epsilon - \bar{x}\|^2] + l\sigma\epsilon^2 && \text{by Theorem 1} \\
&\leq f(x_\epsilon, y(x_\epsilon)) + \mu\epsilon^2 + \frac{1}{2}l'(1+\alpha^2)\epsilon^2 + l\sigma\epsilon^2 && \text{by Lemma 2} \\
&= f(x_\epsilon, y(x_\epsilon)) + \tau\epsilon^2.
\end{aligned}$$

□

Theorem 3 *If the function $\tilde{f}$ admits a feasible descent direction at $\bar{x}$, then $\tilde{f}(\bar{x}_\epsilon) < \tilde{f}(\bar{x})$ for sufficiently small ϵ.*

Proof. Let d ($\|d\| = 1$) be a feasible descent direction at $\bar{x}$ and $\bar{t}$, θ be positive numbers such that

$$\tilde{f}(\bar{x} + td) \leq \tilde{f}(\bar{x}) - \theta t \qquad \text{for all } t \in [0, \bar{t}].$$

Let $\epsilon \in (0, \min\{\bar{t}, \theta/\tau\})$. We have, by Lemma 4:

$$\begin{aligned}
\tilde{f}(\bar{x}_\epsilon) &\leq \tilde{f}(x_\epsilon) + \tau\epsilon^2 \\
&\leq \tilde{f}(\bar{x} + \epsilon d) + \tau\epsilon^2 \\
&\leq \tilde{f}(\bar{x}) - \theta\epsilon + \tau\epsilon^2 \\
&= \tilde{f}(\bar{x}) - \epsilon(\theta - \tau\epsilon) \\
&< \tilde{f}(\bar{x}).
\end{aligned}$$

□

Theorem 4 *Let $\tilde{f}$ admit a feasible descent direction at $\tilde{x}$. Then there exists a positive number ϵ such that $\tilde{f}(\bar{x}_{\bar{\epsilon}}) < \tilde{f}(\tilde{x})$ whenever $\|\tilde{x} - \bar{x}\| \leq \epsilon/2$ and $\bar{\epsilon} \in [\epsilon, 2\epsilon]$.*

Proof. As before, there exist positive numbers $\bar{t}$, θ and a feasible direction d such that $\|d\| = 1$ and

$$\tilde{f}(\tilde{x} + td) \leq \tilde{f}(\tilde{x}) - \theta t \qquad \text{for all} \quad t \in [0, \bar{t}].$$

Let $\epsilon \in (0, \min\{\bar{t}, \theta/8\tau\})$ and $\delta = \epsilon/2$. We have that $\tilde{x} + \delta d$ lies in the ball (trust region) of radius ϵ centered at $\bar{x}$ and:

$$\tilde{f}(\tilde{x} + \delta d) \leq \tilde{f}(\tilde{x}) - \delta\theta.$$

Since $x_{\bar{\epsilon}}$ is optimal within the trust region:

$$\tilde{f}(x_{\bar{\epsilon}}) \leq \tilde{f}(\tilde{x} + \delta d) \leq \tilde{f}(\tilde{x}) - \delta\theta.$$

Now

$$\begin{aligned}
\tilde{f}(\bar{x}_{\bar{\epsilon}}) &\leq \tilde{f}(x_{\bar{\epsilon}}) + \tau\bar{\epsilon}^2 && \text{by Lemma 4} \\
&\leq \tilde{f}(\tilde{x}) - \delta\theta + \tau\bar{\epsilon}^2 \\
&= \tilde{f}(\tilde{x}) - \epsilon(\theta/2 - 4\tau\epsilon) \\
&< \tilde{f}(\tilde{x}).
\end{aligned}$$

□

4 Convergence analysis

We say that a feasible point x is **strongly stationary** for a function g if there do not exist a positive number α and a sequence $\{x^i\}_{i\in I}$ converging to x such that

$$g(x^i) < g(x) - \alpha\|x - x^i\| \qquad \forall i \in I.$$

Note that the set of strong stationary points may be a strict subset of points that are stationary in the Clarke sense, i.e., points for which the zero vector belongs to the generalized subdifferential. Whenever the function g is directionally differentiable over its domain, the set of strong stationary points coincides with the set of B-stationary points, i.e., the set of points where no feasible descent direction exists. In view of Theorem 4 this is the case for the function $f(x, y(x))$ of our bilevel program.

If algorithm BlTrust stops finitely then, since the directional derivatives of the model agree with those of the original objective function, it must do so at a B-stationary point, i.e., a strong stationary point. The next theorem shows that a similar result holds for any accumulation point generated by the trust region procedure.

Theorem 5 *Let $\{x^k\}_{k\in N}$ be an infinite sequence generated by algorithm BlTrust. Then any accumulation point x^* of the sequence is B-stationary, i.e, strongly stationary.*

Proof. Let $\{x^{k_i}\}_{i\in I} \to x^*$. We consider two mutually exclusive cases.

Case 1. There exists a positive number μ such that $\epsilon_{k_i} \geq \mu$ for all $i \in I$.

For this case, we adapt the arguments of Scholtes and Stöhr [10]. Let y^i denote the solution of the lower level linearized variational inequality corresponding to the upper level vector x^i obtained at iteration i and set $f_i(x) = f(x, y^i(x))$. From the statement of the algorithm, it must occur infinitely often that the ratio r_k is larger than or equal to 1/3, i.e., using the previous notation:

$$\tilde{f}(x^{k_i}) - \tilde{f}(x^{k_i+1}) > \frac{1}{3}(\bar{f}(x^{k_i}) - \bar{f}(\bar{x}_{\epsilon_{k_i}})). \tag{1}$$

If x^* is not B-stationary for $\tilde{f}$, then it is not B-stationary either for the function $f_*(x)$ defined as $f_*(x) = f(x, y_*(x))$ where $y_*(x)$ satisfies the variational inequality

$$\langle F(x^*, y^*) + F'_x(x^*, y^*)(x - x^*) + F'_y(x^*, y^*)(y_*(x) - y^*), y_*(x) - y\rangle \leq 0 \qquad \forall y \in Y(x).$$

Hence, for an arbitrary $\epsilon \in (0, \mu]$, one can find a point $\hat{x} \in R^n$ with $\|\hat{x} - x^*\| \leq \epsilon$ such that

$$f(\hat{x}, y_*(\hat{x})) < f(x^*, y(x^*)).$$

From the assumptions and previous continuity results, we obtain that $\|y_*(x) - y_{k_i}(x)\| \to 0$ and $f(\hat{x}, y_{k_i}(\hat{x})) \to f(\hat{x}, y_*(\hat{x}))$. Furthermore:

$$\begin{aligned} f_{k_i}(x^{k_i}) - f_{k_i}(\bar{x}_{\epsilon_{k_i}}) &\geq f_{k_i}(x^{k_i}) - f_{k_i}(\hat{x}) \\ &\geq \frac{1}{2}(f(x^*, y^*) - f(\hat{x}, y_*(\hat{x})) \end{aligned} \tag{2}$$

for every sufficiently large $i \in N$. This implies, in view of (1) and (2), that the series $\sum(\tilde{f}(x^{k_i}) - \tilde{f}(x^{k_i+1}))$ diverges. However, since the trust region method generates a decreasing sequence, we have

$$\sum_{i=1}^{\infty}(\tilde{f}(x^{k_i}) - \tilde{f}(x^{k_i+1})]) \leq \sum_{i=1}^{\infty}(\tilde{f}(x^{k_i}) - \tilde{f}(x^{k_{i+1}})) = \tilde{f}(x^{k_1}) - \tilde{f}(x^*) < \infty,$$

which is the desired contradiction to our assumption that x^* was not B-stationary.

Case 2. $\lim_{i\to\infty} \epsilon_{k_i} = 0$.

Assume that the limit point x^* is not B-stationary and select ϵ as in Theorem 4. Let i be an index such that $\|x^{k_i} - x^*\| \leq \epsilon/2$, j an integer satisfying the relationship $\epsilon \leq 2^j \epsilon_{k_i} = \epsilon' \leq 2\epsilon$. From Theorem 4, we have that $\tilde{f}(\bar{x}_{\epsilon'}) < \tilde{f}(x^*)$ which contradicts the statement that x^* was not a B-stationary limit point of the sequence $\{x^{k_i}\}$. □

Theorem 5 completes our analysis of an approximation scheme for solving bilevel programs. It examplifies the power of a nonconvex approximation for solving nonconvex programs. In our case, this scheme concentrates the combinatorial difficulties on the linearized bilevel program. This decoupling of the combinatorial and nonlinear nature of the problem allows to take full advantage of both the trust region methodology for the outer problem and of efficient combinatorial techniques for the linearized subproblem. This makes for an algorithm that possesses good convergence properties and is implementable. An apparent drawback of the method is the need to perform a costly linesearch at each iteration in order to enforce convergence to a strong stationary point. However this technique will most likely be required only in extreme cases where the algorithm gets trapped in the vicinity of a C-stationary point, an unlikely event.

5 Conclusion

Although algorithm BlTrust is only guaranteed to generate B-stationary limit points, it has the capability to move away from nonglobal stationary points. In this respect, it is unlikely to get trapped into "weak" local solutions and holds the potential of solving to global optimality generalized bilevel programs involving mildly nonlinear functions f and F. For that reason we refer to BlTrust as a "semi-global" method.

Future work on BlTrust will focus on its implementation and on relaxing some of the assumptions required to prove convergence to a B-stationary point. In particular, the set Y could depend nonlinearly on the upper-level vector x. Standard regularity assumptions on this set would ensure that the approximations obtained in our lemmas still hold.

References

[1] Clarke, F. H., *Optimization and Nonsmooth Analysis*, SIAM (1990).

[2] Dafermos, S., "Sensitivity analysis in variational inequalities", *Mathematics of Operations Research* **13** (1988) 421-434.

[3] Fukushima, M., Tseng, P., "An implementable Active-Set algorithm for computing a B-Stationary point of the Mathematical Program with Linear Complementarity Constraints", Technical Report 99020, Dept. of Appl. Math. and Physics, Kyoto University, Oct. 1999.

[4] Hansen, P., Jaumard, B. and Savard, G., "New branch-and-bound rules for linear bilevel programming", *SIAM J. Sci. Stat. Comput.* **13** (1992) 1194-1217.

[5] Luo, Z.Q, Pang, J.S and Ralph, D., *Mathematical programs with equilibrium constraints.* Cambridge University Press (1996).

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

[7] Outrata, J.V., Kočvara, M. and Zowe, J., *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints.* Kluwer Academic Publishers, Dordrecht (1998).

[8] Qiu, Y. and Magnanti, T. L., "Sensitivity analysis for variational inequalities defined on polyhedral sets", *Mathematics of Operations Research* **14** (1989) 410-423.

[9] Robinson, S. M., "Some continuity properties of polyhedral multifunctions" *Mathematical Programming Study* **14** (1981) 206-214.

[10] Scholtes, S. and Stöhr, M., "Exact penalization of mathematical programs with equilibrium constraints" *SIAM J. Control and Optimization* **37** (1999) 617-652.