An Implementable Active-Set Algorithm for Computing a B-Stationary Point of the Mathematical Program with Linear Complementarity Constraints[1]

October 25, 1999

Masao Fukushima[2] and Paul Tseng[3]

Abstract

We consider a mathematical program with smooth objective function and linear inequality/complementarity constraints. We propose an ϵ-active set algorithm which, under a uniform LICQ on the ϵ-feasible set, generates iterates whose cluster points are B-stationary points of the problem. If the objective function is quadratic and ϵ is set to zero, the algorithm terminates finitely. Some numerical experience with the algorithm is reported.

Key words. MPEC, B-stationary point, ϵ-active set.

1 Introduction

We consider the following mathematical program with equilibrium constraints (MPEC):

$$\begin{array}{lll} \text{minimize} & f(z) & \\ \text{subject to} & G_i(z) \geq 0 & i = 1, \ldots, m, \\ & H_i(z) \geq 0 & i = 1, \ldots, m, \\ & G_i(z)H_i(z) = 0 & i = 1, \ldots, m, \\ & g_j(z) \leq 0 & j = 1, \ldots, p, \\ & h_l(z) = 0 & l = 1, \ldots, q, \end{array} \tag{1}$$

where f is a real-valued continuously differentiable function on $\Re^n$ and G_i, H_i, g_j, h_l are real-valued affine functions on $\Re^n$.

This problem has been of much interest and many algorithms have been proposed for its solution, as is evidenced by [1, 2, 5, 7, 9] and the extensive references therein. However, these algorithms in general are only guaranteed to compute either a B-stationary point under the nondegeneracy (strict complementarity) assumption that is somewhat restrictive in practice, or a C-stationary point for the problem, rather than the desired B-stationary point. Scholtes and Stöhr [12] showed that a trust region method converges to a B-stationary point provided the trust region radii do not tend to zero. However, it is not clear whether we can expect the latter condition to hold in the degenerate case. Recently, Fukushima and Pang [3] considered a continuation method based on a smoothed

[1] This research is supported by Scientific Research Grant-in-Aid from the Ministry of Education, Science, Sports and Culture of Japan, and National Science Foundation Grant CCR-9731273.

[2] Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan. (fuku@kuamp.kyoto-u.ac.jp)

[3] Department of Mathematics, University of Washington, Seattle, Washington 98195, U.S.A. (tseng@math.washington.edu)

approximation of MPEC and showed, under LICQ (for MPEC) and an "asymptotic weak nondegeneracy" assumption, convergence of second-order stationary point of the smoothed problem to a B-stationary point of MPEC as the smoothing parameter tends to zero. Subsequently, Scholtes [11] showed an analogous result for a method based on another smoothed approximation of MPEC. However, the algorithms considered in [3, 11] are conceptual, and global convergence to a B-stationary point has yet to be established for an implementable algorithm for solving MPECs.

The purpose of this paper is to propose an implementable ϵ-active set algorithm for solving MPEC (1) and establish global convergence of the proposed algorithm to a B-stationary point of (1) under a uniform LICQ on the ϵ-feasible set. Moreover, we show that if the objective function is quadratic and ϵ is set to zero, then the algorithm terminates finitely at a B-stationary point of (1).

A few words about notation: Since functions G_i, H_i, g_j, and h_l are all assumed to be affine, the gradients of these functions are constant vectors. Nevertheless, we will throughout write $\nabla G_i(z^k)$, $\nabla H_i(z^k)$, etc., to specify the point under consideration. This will better illustrate the nature of the proposed algorithm and might suggest possible extension to problems involving nonlinear constraints.

2 Some Properties

Let $\mathcal{P}$ denote the set of all pairs (A, B) such that A and B are subsets of $\{1, \ldots, m\}$ and $A \cup B = \{1, \ldots, m\}$. Fix $\epsilon \geq 0$. For each $(A, B) \in \mathcal{P}$, define the restricted ϵ-feasible set:

$$
\begin{aligned}
\mathcal{F}_\epsilon[A, B] \;:=\; \{\, z \in \Re^n : \;& 0 \leq G_i(z) \leq \epsilon \quad i \in A, \\
& 0 \leq G_i(z) \qquad\;\; i \notin A, \\
& 0 \leq H_i(z) \leq \epsilon \quad i \in B, \\
& 0 \leq H_i(z) \qquad\;\; i \notin B, \\
& g_j(z) \leq 0 \qquad\;\;\; j = 1, \ldots, p, \\
& h_l(z) = 0 \qquad\;\;\; l = 1, \ldots, q \,\}.
\end{aligned}
$$

Note that the sets $\{\mathcal{F}_\epsilon[A, B] : (A, B) \in \mathcal{P}\}$ are not mutually disjoint. Define the ϵ-feasible set for MPEC (1):

$$\mathcal{F}_\epsilon := \bigcup_{(A,B)\in\mathcal{P}} \mathcal{F}_\epsilon[A, B].$$

Then, for $\epsilon = 0$, $\mathcal{F}_\epsilon$ is just the feasible set for MPEC (1).

For any feasible solution $\bar{z} \in \mathcal{F}_0$ of MPEC (1), let $\mathcal{T}(\bar{z}; \mathcal{F}_0)$ denote the tangent cone of $\mathcal{F}_0$ at $\bar{z}$. Then $\bar{z}$ is called a *B-stationary point* of MPEC (1) if it satisfies

$$\nabla f(\bar{z})^T d \geq 0 \quad \forall\, d \in \mathcal{T}(\bar{z}; \mathcal{F}_0).$$

In MPEC, the tangent cone $\mathcal{T}(\bar{z}; \mathcal{F}_0)$ is normally represented as a finite union of closed convex cones [7] and hence, in general, nonconvex, unless the nondegeneracy (strict complementarity) condition is satisfied. This fact gives rise to a combinatorial nature of MPEC that makes a problem intractable. Recent attempts to identify a favorable class of

MPECs have been focused on constraint qualifications pertaining to MPEC which enable us to characterize a B-stationary point in a simple and convenient manner [10, 13].

For each $z \in \mathcal{F}_\epsilon$, define the index sets:

$$
\begin{aligned}
A_\epsilon(z) &:= \{i : G_i(z) \le \epsilon\}, \\
B_\epsilon(z) &:= \{i : H_i(z) \le \epsilon\}, \\
I_\epsilon(z) &:= \{j : g_j(z) \ge -\epsilon\}.
\end{aligned}
$$

By the definition of $\mathcal{P}$, we always have $A_\epsilon(z) \cup B_\epsilon(z) = \{1, \ldots, m\}$. Associated with any feasible solution $\bar{z}$ of MPEC (1) is the following *relaxed problem*, denoted $\mathrm{R}(\bar{z})$:

$$
\begin{array}{lll}
\text{minimize} & f(z) & \\
\text{subject to} & G_i(z) = 0 & i \in A_0(\bar{z}) \setminus B_0(\bar{z}), \\
& G_i(z) \ge 0 & i \in B_0(\bar{z}), \\
& H_i(z) = 0 & i \in B_0(\bar{z}) \setminus A_0(\bar{z}), \\
& H_i(z) \ge 0 & i \in A_0(\bar{z}), \\
& g_j(z) \le 0 & j = 1, \ldots, p, \\
& h_l(z) = 0 & l = 1, \ldots, q.
\end{array} \tag{2}
$$

Note that the relaxed problem (2) is an ordinary nonlinear program with linear constraints.

We assume the following *uniform LICQ* on $\mathcal{F}_\epsilon$ for some $\epsilon \ge 0$:

$$
\eta \;<\; \left\| \sum_{i \in A_\epsilon(z)} \nabla G_i(z) v_i + \sum_{i \in B_\epsilon(z)} \nabla H_i(z) w_i - \sum_{j \in I_\epsilon(z)} \nabla g_j(z) \lambda_j - \sum_{l=1}^{q} \nabla h_l(z) \mu_l \right\|_1
$$

$$
\text{whenever} \quad z \in \mathcal{F}_\epsilon \quad \text{and} \quad \max_{i,j,l} \{|v_i|, |w_i|, |\lambda_j|, |\mu_l|\} > 1,
$$

where $\eta > 0$ is some constant. Notice that since G_i, H_i, g_j, h_l are affine for all i, j, l, we can equivalently replace η by 0 in this assumption. However, the constant η will play a useful role in our algorithm and its analysis. Recall that $\bar{z} \in \mathcal{F}_0$ implies $A_0(\bar{z}) \cup B_0(\bar{z}) = \{1, \ldots, m\}$. Hence, for $\epsilon = 0$, the uniform LICQ on $\mathcal{F}_0$ essentially amounts to LICQ (in the sense of ordinary nonlinear programming) for the relaxed problem $\mathrm{R}(\bar{z})$ being satisfied at every feasible solution $\bar{z}$ of MPEC (1). The following theorem will play an essential role in the subsequent analysis.

Theorem 1 *Let $\bar{z}$ be a feasible solution of MPEC (1) such that the LICQ holds for the relaxed problem $R(\bar{z})$. If $\bar{z}$ is a KKT point of the relaxed problem $R(\bar{z})$, then $\bar{z}$ is a B-stationary point of MPEC (1).*

This theorem has been proved under a more general setting in [10] (see also [7, Proposition 4.3.7] and [13] for related results). The significance of the result stated in Theorem 1 lies in the fact that, under the LICQ for the relaxed problem $\mathrm{R}(\bar{z})$, B-stationarity for MPEC can be completely characterized by the KKT conditions for problem $\mathrm{R}(\bar{z})$, which is an ordinary nonlinear program. This observation has paved the way to developing iterative methods that generate a sequence converging to a B-stationary point of MPEC under the LICQ for the relaxed problem [3, 11].

3 An ϵ-Active Set Algorithm

Consider the following ϵ-active set algorithm for MPEC (1): Choose arbitrary constants $\sigma_1 \in (0,1), \sigma_2 \in (0,1)$, $t_{\max} \in (0,\infty)$, $\omega_1 \in (0,1)$, $\omega_2 \in [0,1)$, and a continuous function $\rho : \Re_+ \to \Re_+$ such that $\rho(\nu) = 0 \;\Leftrightarrow\; \nu = 0$. At each iteration $k = 0,1,\ldots$, we begin with some $\epsilon_k \in [0,\epsilon]$ and $\nu_k \geq 0$, some $(A^k, B^k) \in \mathcal{P}$ and some $z^k \in \mathcal{F}_{\epsilon_k}[A^k, B^k]$. We consider the following subproblem associated with the current iterate z^k and index set pair (A^k, B^k):

$$
\begin{array}{lll}
\text{minimize} & f(z) & \\
\text{subject to} & G_i(z) = G_i(z^k) & i \in A^k, \\
& G_i(z) \geq 0 & i \in B^k \setminus A^k, \\
& H_i(z) = H_i(z^k) & i \in B^k, \\
& H_i(z) \geq 0 & i \in A^k \setminus B^k, \\
& g_j(z) \leq 0 & j = 1,\ldots,p, \\
& h_l(z) = 0 & l = 1,\ldots,q.
\end{array}
\tag{3}
$$

Notice that (3) is a linearly constrained problem with z^k as a feasible solution, so we can solve this problem (approximately) by using a feasible descent algorithm with starting point z^k. This is described in Step 1 below.

Step 1: We compute a $\hat{z}^k \in \mathcal{F}_{\epsilon_k}[A^k, B^k]$ satisfying

$$f(\hat{z}^k) \leq f(z^k), \tag{4}$$

and, for some $\delta_k \in [0, \eta\nu_k/2]$, r^k, $(v_i^k)_{i\in\hat{A}^k}$, $(w_i^k)_{i\in\hat{B}^k}$, $(\lambda_j^k)_{j\in I^k}$, $(\mu_l^k)_{l=1}^q$,

$$
\begin{aligned}
-r^k &= \nabla f(\hat{z}^k) - \sum_{i\in\hat{A}^k} \nabla G_i(\hat{z}^k) v_i^k - \sum_{i\in\hat{B}^k} \nabla H_i(\hat{z}^k) w_i^k \\
&\quad + \sum_{j\in I^k} \nabla g_j(\hat{z}^k)\lambda_j^k + \sum_{l=1}^{q} \nabla h_l(\hat{z}^k)\mu_l^k, \\
&\|r^k\|_1 \leq \delta_k, \\
&v_i^k \geq 0 \quad \forall i \in \hat{A}^k \setminus A^k, \\
&w_i^k \geq 0 \quad \forall i \in \hat{B}^k \setminus B^k, \\
&\lambda_j^k \geq 0 \quad \forall j \in I^k,
\end{aligned}
\tag{5}
$$

where e is the vector of 1s and

$$\hat{A}^k := A_{\epsilon_k}(\hat{z}^k),\;\; \hat{B}^k := B_{\epsilon_k}(\hat{z}^k),\;\; I^k := I_{\epsilon_k}(\hat{z}^k).$$

Roughly speaking, $\hat{z}^k$ is an approximate KKT point of the subproblem (3). Moreover we have from $\hat{A}^k \supseteq A^k$, $\hat{B}^k \supseteq B^k$ and $A^k \cup B^k = \{1,\ldots,m\}$ that

$$
\begin{aligned}
\hat{A}^k \setminus \hat{B}^k &= A^k \setminus \hat{B}^k \subseteq A^k \setminus B^k, \\
\hat{B}^k \setminus \hat{A}^k &= B^k \setminus \hat{A}^k \subseteq B^k \setminus A^k, \\
\hat{A}^k \cap \hat{B}^k &\supseteq A^k \cap B^k.
\end{aligned}
$$

[**Comment**: See Note 1 below on computing $\hat{z}^k$.]

Step 2: We have either (i) $\hat{A}^k \cap \hat{B}^k$ is a proper superset of $A^k \cap B^k$ or (ii) $\hat{A}^k \cap \hat{B}^k = A^k \cap B^k$. In case (i), set

$$\nu_{k+1} := \nu_k, \quad \epsilon_{k+1} := \epsilon_k,$$
$$z^{k+1} := \tilde{z}^k := \hat{z}^k,$$
$$(A^{k+1}, B^{k+1}) := (\hat{A}^k, \hat{B}^k)$$

and exit. In case (ii), we proceed to Step 3. [**Comment**: Notice that case (i) can occur in at most $m-1$ consecutive iterations.]

Step 3: We have $\hat{A}^k \setminus \hat{B}^k = A^k \setminus B^k$ and $\hat{B}^k \setminus \hat{A}^k = B^k \setminus A^k$, implying $\hat{A}^k = A^k$ and $\hat{B}^k = B^k$. If $v_i^k \geq -\nu_k$ and $w_i^k \geq -\nu_k$ for all $i \in A^k \cap B^k$, then we set

$$\nu_{k+1} := \omega_1 \nu_k, \quad \tilde{z}^k := \hat{z}^k,$$

and proceed to Step 5. Otherwise, we proceed to Step 4.

Step 4: There exists some index $i_k \in A^k \cap B^k$ such that either (iia) $v_{i_k}^k < -\nu_k$ or (iib) $w_{i_k}^k < -\nu_k$. Then, by the uniform LICQ, the following linear system has a solution d^k:

$$\begin{array}{rcll}
\nabla f(\hat{z}^k)^T d & \leq & -\eta \nu_k/2, & \\
-e \;\leq\; d & \leq & e, & \\
\nabla G_i(\hat{z}^k)^T d & = & 0 & \forall i \in A^k \setminus \{i_k\}, \\
\nabla H_i(\hat{z}^k)^T d & = & 0 & \forall i \in B^k \setminus \{i_k\}, \\
\nabla G_{i_k}(\hat{z}^k)^T d & \geq & 0 \quad \text{in case (iia)}, & \\
\nabla H_{i_k}(\hat{z}^k)^T d & = & 0 \quad \text{in case (iia)}, & \\
\nabla H_{i_k}(\hat{z}^k)^T d & \geq & 0 \quad \text{in case (iib)}, & \\
\nabla G_{i_k}(\hat{z}^k)^T d & = & 0 \quad \text{in case (iib)}, & \\
\nabla G_i(\hat{z}^k)^T d & \geq & -G_i(\hat{z}^k) & \forall i \in A_\epsilon(\hat{z}^k) \setminus A^k, \\
\nabla H_i(\hat{z}^k)^T d & \geq & -H_i(\hat{z}^k) & \forall i \in B_\epsilon(\hat{z}^k) \setminus B^k, \\
\nabla g_j(\hat{z}^k)^T d & \leq & 0 & \forall j \in I^k, \\
\nabla g_j(\hat{z}^k)^T d & \leq & -g_j(\hat{z}^k) & \forall j \in I_\epsilon(\hat{z}^k) \setminus I^k, \\
\nabla h_l(\hat{z}^k)^T d & = & 0 & l = 1, \ldots, q
\end{array} \tag{6}$$

(with "$\leq -\eta\nu_k/2$" replaced by "< 0" in the case of $\nu_k = 0$). [**Comment**: See Note 2 below for a justification of this fact.] Let $\bar{t}_k := \min\{t_{\max}, \sup\{t : \hat{z}^k + td^k \in \mathcal{F}_{\epsilon_k}\}\}$, and set

$$\nu_{k+1} := \nu_k, \quad \tilde{z}^k := \hat{z}^k + t_k d^k,$$

where t_k is the largest element of $\{\bar{t}_k, \sigma_1 \bar{t}_k, \sigma_1^2 \bar{t}_k, \ldots\}$ satisfying

$$f(\hat{z}^k + td^k) \leq f(\hat{z}^k) + \sigma_2 t \nabla f(\hat{z}^k)^T d^k.$$

[**Comment**: Notice that $\bar{t}_k$ is computable by a minimum ratio formula. Moreover, we have $\bar{t}_k \geq \min\{t_{\max}, \epsilon_k / \max_{i,j}\{\|\nabla G_i(\hat{z}^k)\|_1, \|\nabla H_i(\hat{z}^k)\|_1, \|\nabla g_j(\hat{z}^k)\|_1\}\}$ and, by a standard argument for Armijo stepsize rule, $t_k > 0$.]

Step 5: Let $\tilde{A}^k := A_{\epsilon_k}(\tilde{z}^k)$, $\tilde{B}^k := B_{\epsilon_k}(\tilde{z}^k)$. Set

$$\begin{aligned}\epsilon_{k+1} &:= \min\{\max\{\rho(\nu_{k+1}), \omega_2\epsilon_k\}, \epsilon_k\} \;=\; \mathrm{mid}\{\omega_2\epsilon_k, \rho(\nu_{k+1}), \epsilon_k\},\\ z^{k+1} &:= \arg\min_{z\in\mathcal{F}_{\epsilon_{k+1}}[\tilde{A}^k,\tilde{B}^k]} \|z-\tilde{z}^k\|,\\ (A^{k+1},B^{k+1}) &:= (A_{\epsilon_{k+1}}(z^{k+1}), B_{\epsilon_{k+1}}(z^{k+1}))\end{aligned}$$

and exit.

Note 1: In view of the relations

$$\hat{A}^k = A^k \cup (\hat{A}^k \setminus A^k), \quad \hat{B}^k = B^k \cup (\hat{B}^k \setminus B^k),$$

we remark that $\hat{z}^k$ satisfies (5) whenever the following linear system has no solution:

$$\begin{aligned}&\nabla f(\hat{z}^k)^T d \le -\delta_k,\\ &-e \;\le\; d \;\le\; e,\\ &\nabla G_i(\hat{z}^k)^T d = 0 \quad \forall\, i \in A^k,\\ &\nabla G_i(\hat{z}^k)^T d \ge 0 \quad \forall\, i \in \hat{A}^k \setminus A^k,\\ &\nabla H_i(\hat{z}^k)^T d = 0 \quad \forall\, i \in B^k,\\ &\nabla H_i(\hat{z}^k)^T d \ge 0 \quad \forall\, i \in \hat{B}^k \setminus B^k,\\ &\nabla g_j(\hat{z}^k)^T d \le 0 \quad \forall j \in I^k,\\ &\nabla h_l(\hat{z}^k)^T d = 0 \quad \forall\, l = 1,\dots,q\end{aligned} \tag{7}$$

(with "$\le -\delta_k$" replaced by "< 0" whenever $\delta_k = 0$). This is because, by Farkas lemma, the system (7) has no solution if and only if the following dual linear system has a solution:

$$\begin{aligned}-1 \;&\ge\; -\delta_k\lambda_0 + e^T\pi^+ + e^T\pi^-,\\ 0 \;=\; \nabla f(\hat{z}^k)\lambda_0 &+ \pi^+ - \pi^- - \sum_{i\in\hat{A}^k} \nabla G_i(\hat{z}^k)v_i - \sum_{i\in\hat{B}^k} \nabla H_i(\hat{z}^k)w_i\\ &+ \sum_{j\in I^k} \nabla g_j(\hat{z}^k)\lambda_j + \sum_{l=1}^{q} \nabla h_l(\hat{z}^k)\mu_l,\\ &\lambda_0 \ge 0, \quad \pi^+ \ge 0, \quad \pi^- \ge 0,\\ &\lambda_j \ge 0 \quad \forall\, j \in I^k,\\ &v_i \ge 0 \quad \forall\, i \in \hat{A}^k\setminus A^k,\\ &w_i \ge 0 \quad \forall\, i \in \hat{B}^k\setminus B^k.\end{aligned}$$

The first inequality implies that $\lambda_0 > 0$ and that $(e^T\pi^+ + e^T\pi^-)/\lambda_0 \le \delta_k$. Then, dividing the above system by λ_0 and setting $r^k = (\pi^+ - \pi^-)/\lambda_0$ and $v_i^k = v_i/\lambda_0$, $w_i^k = w_i/\lambda_0$, $\lambda_j^k = \lambda_j/\lambda_0$, $\mu_l^k = \mu_l/\lambda_0$, we obtain (5), because

$$\|r^k\|_1 = \|\pi^+ - \pi^-\|_1/\lambda_0 \le (\|\pi^+\|_1 + \|\pi^-\|_1)/\lambda_0 = (e^T\pi^+ + e^T\pi^-)/\lambda_0 \le \delta_k.$$

For the case of $\epsilon_k > 0$ and $\nu_k > 0$, a point $\hat{z}^k$ such that (4) holds and (7) has no solution is computable in finite time by any feasible descent method for the linearly constrained subproblem (3), with starting point z^k, whose generated points converge in subsequence to a KKT point of (3). [It would be enough to find a point $\hat{z}^k$ sufficiently near to a KKT point z of problem (3) so that $A_0(z) \subseteq \hat{A}^k$, $B_0(z) \subseteq \hat{B}^k$, $I_0(z) \subseteq I^k$, and $\|\nabla f(\hat{z}^k) - \nabla f(z)\|_1 < \delta_k$. In fact, for such $\hat{z}^k$, the system (7) has no solution, because any solution of (7) would be a feasible descent direction for problem (3) at z, contradicting z being a KKT point of (3).] For the case of $\epsilon_k = 0$ and $\nu_k = 0$ and f is quadratic, such a $\hat{z}^k$ is a KKT point of (3) and is computable in finite time by using a feasible descent method coupled with an active-set identification strategy. Notice that (3) need not have a unique KKT point.

Note 2: Suppose the system (6), with "$\leq -\eta\nu_k/2$" replaced by "< 0" in the case of $\nu_k = 0$, does not have a solution. Then, by Farkas lemma, its dual system has a solution so that, by the same argument as in Note 1, and using the fact that, in this case,

$$A^k = \hat{A}^k = A_{\epsilon_k}(\hat{z}^k) \subseteq A_\epsilon(\hat{z}^k),$$
$$B^k = \hat{B}^k = B_{\epsilon_k}(\hat{z}^k) \subseteq B_\epsilon(\hat{z}^k),$$
$$I^k = I_{\epsilon_k}(\hat{z}^k) \subseteq I_\epsilon(\hat{z}^k),$$

the following linear system has a solution:

$$\begin{aligned}
&\lambda_j \geq 0 \quad \forall\, j \in I_\epsilon(\hat{z}^k),\\
&v_{i_k} \geq 0 \ \text{ in case (iia)}, \quad w_{i_k} \geq 0 \ \text{ in case (iib)},\\
&v_i \geq 0 \quad \forall\, i \in A_\epsilon(\hat{z}^k) \setminus A^k,\\
&w_i \geq 0 \quad \forall\, i \in B_\epsilon(\hat{z}^k) \setminus B^k,\\
-r \;=\; &\nabla f(\hat{z}^k) - \sum_{i \in A_\epsilon(\hat{z}^k)} \nabla G_i(\hat{z}^k) v_i - \sum_{i \in B_\epsilon(\hat{z}^k)} \nabla H_i(\hat{z}^k) w_i\\
&+ \sum_{j \in I_\epsilon(\hat{z}^k)} \nabla g_j(\hat{z}^k)\lambda_j + \sum_{l=1}^{q} \nabla h_l(\hat{z}^k)\mu_l,
\end{aligned}$$

where $\|r\|_1 \leq \eta\nu_k/2$. Subtracting the equation from that in (5) yields

$$\begin{aligned}
r - r^k \;=\; &- \sum_{i \in A_\epsilon(\hat{z}^k)} \nabla G_i(\hat{z}^k)(v_i^k - v_i) - \sum_{i \in B_\epsilon(\hat{z}^k)} \nabla H_i(\hat{z}^k)(w_i^k - w_i)\\
&+ \sum_{j \in I_\epsilon(\hat{z}^k)} \nabla g_j(\hat{z}^k)(\lambda_j^k - \lambda_j) + \sum_{l=1}^{q} \nabla h_l(\hat{z}^k)(\mu_l^k - \mu_l),
\end{aligned}$$

with v_i^k (respectively, w_i^k, λ_j^k) set to zero if it is not defined in (5). Since $v_{i_k} - v_{i_k}^k > \nu_k$ in case (iia) and $w_{i_k} - w_{i_k}^k > \nu_k$ in case (iib), while $\delta_k \leq \eta\nu_k/2$ so that $\|r^k\|_1 \leq \delta_k \leq \eta\nu_k/2$, we would have $\|r - r^k\|_1 \leq \|r\|_1 + \|r^k\|_1 \leq \eta\nu_k$, contradicting the uniform LICQ at $\hat{z}^k$. Thus, the system (6) has a solution.

Note 3: In the 5th, 6th and 8th inequalities of (6), the right-hand side can alternatively be replaced by zero. However, the resulting linear system would have a smaller solution set.

Note 4: As described, the algorithm requires knowledge of η in choosing δ_k. If η is not known, we can estimate η on-line using the following back-tracking scheme: In Step 4, if (iia) or (iib) occurs but (6) does not have a solution, then decrease η by a constant fraction and repeat iteration k. [The initial η is chosen arbitrarily.] Under the uniform LICQ, η would be decreased only a finite number of times (see Note 2), and the convergence result given in Theorem 2 would still hold for this variant of the algorithm.

Note 5: In Step 3, instead of setting ν_{k+1} to $\omega_1\nu_k$, one can more generally set ν_{k+1} to be the term after ν_k along some pre-specified decreasing sequence (not necessarily geometric) tending to zero. The term $\omega_2\epsilon_k$ in Step 5 can be similarly generalized. The updating rule for ϵ_k enforces that ϵ_k goes to zero at the rate of $\rho(\nu_k)$. Thus, ϵ_k, which measures primal feasibility, and ν_k, which measures dual feasibility, can go to zero at different rates, depending on the choice of the function ρ.

4 Convergence to B-Stationary Point

To establish convergence of the preceding algorithm, we assume f is bounded below on $\mathcal{F}_\epsilon$. We further assume $\{z^k\}$ and $\{\hat{z}^k\}$ are bounded (which would be implied by the boundedness of $\mathcal{F}_\epsilon$). A sufficient condition for this assumption to hold is given in Lemma 1 at the end of this section.

Theorem 2 *Under the above assumptions (in addition to the uniform LICQ on $\mathcal{F}_\epsilon$), the following hold for $\{(\epsilon_k, \nu_k, z^k, \hat{z}^k, \tilde{z}^k)\}_{k=0,1,\ldots}$ generated by the ϵ-active set algorithm:*
(a) If $\epsilon_0 > 0$, $\nu_0 > 0$ and f is Lipschitz continuous with constant L on a set Z containing $\{z^k\}$ and $\{\tilde{z}^k\}$, then every cluster point of $\{z^k\}$ is a B-stationary point of MPEC (1).
(b) If $\epsilon_0 = \nu_0 = 0$ and f is quadratic, then there exists $\bar{k} \in \{0,1,\ldots\}$ such that $z^{\bar{k}}$ is a B-stationary point of MPEC (1).

Proof. Let $\mathcal{K} := \{k \in \{0,1,\ldots\}$: Step 5 is entered from Step 4 in iteration $k\}$, $\mathcal{K}' := \{k \in \{0,1,\ldots\}$: Step 5 is entered from Step 3 in iteration $k\}$. Then, we have $z^k \in \mathcal{F}_{\epsilon_k}$ for all k, and

$$f(\tilde{z}^k) \le f(\hat{z}^k) \le f(z^k) \quad \text{and} \quad \hat{z}^k \in \mathcal{F}_{\epsilon_k} \qquad \forall\, k \in \mathcal{K} \cup \mathcal{K}', \tag{8}$$

and $z^{k+1} = z^k$ for all $k \notin \mathcal{K} \cup \mathcal{K}'$.

(a) Suppose $\epsilon_0 > 0, \nu_0 > 0$ and f is Lipschitz continuous with constant L on set Z containing $\{z^k\}$ and $\{\tilde{z}^k\}$. If $\nu_k \to 0$, then $|\mathcal{K}'| = \infty$, $\delta_k \to 0$, and the updating formula for ϵ_k would imply $\epsilon_k \to 0$, so any cluster point $\bar{z}$ of $\{\hat{z}^k\}_{k\in\mathcal{K}'}$ would be a KKT point of the *relaxed problem* $\mathrm{R}(\bar{z})$, which is a B-stationary point of MPEC (1) under the (uniform) LICQ. Suppose instead $\nu_k \not\to 0$, so that $|\mathcal{K}'| < \infty$, $|\mathcal{K}| = \infty$, and $\nu = \lim_{k\to\infty} \nu_k > 0$. We will obtain a contradiction below.

For each $k \in \mathcal{K} \cup \mathcal{K}'$, $\tilde{z}^k$ satisfies all constraints defining $\mathcal{F}_{\epsilon_{k+1}}[\tilde{A}^k, \tilde{B}^k]$, except for possibly

$$G_i(z) \le \epsilon_{k+1} \quad \forall\, i \in \tilde{A}^k,$$
$$H_i(z) \le \epsilon_{k+1} \quad \forall\, i \in \tilde{B}^k.$$

Then, by a well-known lemma of Hoffman [4], there exists a constant $\tau > 0$ (depending on $\nabla G_i, \nabla H_i, \nabla g_j, \nabla h_l$ only) such that

$$\begin{aligned}
\|z^{k+1} - \tilde{z}^k\| &\leq \tau \left(\sum_{i \in \tilde{A}^k} \left| [G_i(\tilde{z}^k) - \epsilon_{k+1}]_+ \right| + \sum_{i \in \tilde{B}^k} \left| [H_i(\tilde{z}^k) - \epsilon_{k+1}]_+ \right| \right) \\
&= \tau \left(\sum_{i \in \tilde{A}^k} \left| [G_i(\tilde{z}^k) - \epsilon_{k+1}]_+ - [G_i(\tilde{z}^k) - \epsilon_k]_+ \right| \right. \\
&\quad \left. + \sum_{i \in \tilde{B}^k} \left| [H_i(\tilde{z}^k) - \epsilon_{k+1}]_+ - [H_i(\tilde{z}^k) - \epsilon_k]_+ \right| \right) \\
&\leq \tau \left(\sum_{i \in \tilde{A}^k} |\epsilon_{k+1} - \epsilon_k| + \sum_{i \in \tilde{B}^k} |\epsilon_{k+1} - \epsilon_k| \right) \\
&\leq 2\tau m(\epsilon_k - \epsilon_{k+1}), \qquad (9)
\end{aligned}$$

where the equality follows from the fact that $\tilde{z}^k \in \mathcal{F}_{\epsilon_k}[\tilde{A}^k, \tilde{B}^k]$ and the second inequality uses the nonexpansive property of $[\cdot]_+ := \max\{0, \cdot\}$ with respect to $|\cdot|$. Since f is Lipschitz continuous with constant L on Z containing z^{k+1} and $\tilde{z}^k$, it follows that

$$f(z^{k+1}) \leq f(\tilde{z}^k) + L\|z^{k+1} - \tilde{z}^k\| \leq f(\tilde{z}^k) + 2L\tau m(\epsilon_k - \epsilon_{k+1}). \qquad (10)$$

This together with (8) yields

$$f(z^{k+1}) \leq f(z^k) + 2L\tau m(\epsilon_k - \epsilon_{k+1}) \quad \forall\, k \in \mathcal{K} \cup \mathcal{K}',$$

and $f(z^{k+1}) = f(z^k)$ for all $k \notin \mathcal{K} \cup \mathcal{K}'$. Since, by our assumption, $\{f(z^k)\}$ is bounded below, and $\{\epsilon_k\}$ is monotonically nonincreasing and positive, this in turn implies that $\{f(z^k)\}$ converges and so $f(z^{k+1}) - f(z^k) \to 0$. Since, for each $k \in \mathcal{K}$,

$$\begin{aligned}
f(\tilde{z}^k) &= f(\hat{z}^k + t_k d^k) \\
&\leq f(\hat{z}^k) + \sigma_2 t_k \nabla f(\hat{z}^k)^T d^k \\
&\leq f(z^k) - \sigma_2 t_k \eta \nu_k / 2 \\
&\leq f(z^k) - \sigma_2 t_k \eta \nu / 2,
\end{aligned}$$

this and (10) imply $\{t_k\}_{k \in \mathcal{K}} \to 0$.

The Armijo-type stepsize rule for determining t_k implies, for each $k \in \mathcal{K}$, either (i) $t_k = \bar{t}_k$ or (ii) $t_k < \bar{t}_k$. Since $\nu = \lim_{k \to \infty} \nu_k > 0$, the updating rule for ϵ_{k+1} implies $\lim_{k \to \infty} \epsilon_k > 0$, so $\bar{t}_k$ is bounded away from zero (see the comment at the end of Step 4 of the algorithm). Since $\{t_k\}_{k \in \mathcal{K}} \to 0$, this implies case (i) can occur for only a finite number of iterations $k \in \mathcal{K}$, so it must be that case (ii) occurs for all $k \in \mathcal{K}$ sufficiently large, in which case

$$f(\hat{z}^k + (t_k/\sigma_1)d^k) - f(\hat{z}^k) > \sigma_2 (t_k/\sigma_1) \nabla f(\hat{z}^k)^T d^k.$$

Since $\{\hat{z}^k\}_{k \in \mathcal{K}}$ is assumed bounded and $-e \leq d^k \leq e$ for all $k \in \mathcal{K}$, this together with $\{t_k\}_{k \in \mathcal{K}} \to 0$ would yield

$$\nabla f(\hat{z}^\infty)^T d^\infty \geq \sigma_2 \nabla f(\hat{z}^\infty)^T d^\infty,$$

where $(\hat{z}^\infty, d^\infty)$ is any cluster point of $\{(\hat{z}^k, d^k)\}_{k\in\mathcal{K}}$. Then, we would have $\nabla f(\hat{z}^\infty)^T d^\infty \geq 0$, contradicting the fact that $\nabla f(\hat{z}^k)^T d^k \leq -\eta\nu_k/2 \leq -\eta\nu/2$ for all $k \in \mathcal{K}$.

(b) Suppose $\epsilon_0 = \nu_0 = 0$ and f is quadratic. In this case, $\epsilon_k = \nu_k = 0$ for all k. Since we can exit from Step 2 in at most $m-1$ consecutive iterations before exiting from Step 5, it suffices to show that there exists $\bar{k} \in \mathcal{K} \cup \mathcal{K}'$ such that $z^{\bar{k}}$ is a B-stationary point of MPEC (1). For each $k \in \mathcal{K}$, we have from $t_k > 0$ and $z^{k+1} = \tilde{z}^k$ that

$$f(z^{k+1}) = f(\hat{z}^k + t_k d^k) \leq f(\hat{z}^k) + \sigma_2 t_k \nabla f(\hat{z}^k)^T d^k < f(\hat{z}^k) \leq f(z^k),$$

so the values $f(\hat{z}^k)$, $k \in \mathcal{K}$, are distinct. On the other hand, $\nu_k = 0$ implies $\delta_k = 0$ for all k , so (5) yields that $\hat{z}^k$ is an (exact) stationary point of the subproblem (3). Since (3) is a quadratic program, a lemma from [8] shows that the set of values

$$\{\ f(z)\ :\ z \text{ is a stationary point of (3)}\ \}$$

is *finite*. Since $\epsilon_k = 0$ for all k and the number of index set pairs $(A, B) \in \mathcal{P}$ is finite, the number of distinct quadratic programs of the form (3) is finite. It follows that the sequence $\{f(\hat{z}^k)\}_{k\in\mathcal{K}}$ is finite. Hence there must be an index $k \in \mathcal{K}'$. If $k \in \mathcal{K}'$, then Step 5 is entered from Step 3 so that $v_i^k \geq -\nu_k = 0$ and $w_i^k \geq -\nu_k = 0$ for all $i \in A^k \cap B^k$. Moreover $\hat{z}^k$ satisfies (5) with $\epsilon_k = \delta_k = 0$. Therefore $\hat{z}^k$ is a KKT point of the relaxed problem $\mathrm{R}(\hat{z}^k)$, and hence a B-stationary point of MPEC (1) under the LICQ for $\mathrm{R}(\hat{z}^k)$. ■

Lemma 1 *Suppose there exists $L > 0$ such that*

$$\tilde{\mathcal{F}}_{\epsilon_0} := \{z \in \mathcal{F}_{\epsilon_0} : f(z) \leq f(z^0) + 2L\tau m\epsilon_0\}$$

is bounded and f is Lipschitz continuous on $Z := \tilde{\mathcal{F}}_{\epsilon_0} + 2\tau m\epsilon_0\mathcal{B}$ with constant L, where $\mathcal{B}$ denotes the unit sphere in $\Re^n$. [Such an L exists when, for example, $\mathcal{F}_{\epsilon_0}$ is bounded.] Then, $\{z^k\}$ and $\{\hat{z}^k\}$ lie in $\tilde{\mathcal{F}}_{\epsilon_0}$ and hence are bounded, and $\{z^k\}$ and $\{\tilde{z}^k\}$ lie in Z.

Proof. An induction argument using (9) and (10) yields, for each $k \in \mathcal{K} \cup \mathcal{K}'$,

$$z^k, \hat{z}^k \in \mathcal{F}_{\epsilon_k}, \quad f(\tilde{z}^k) \leq f(\hat{z}^k) \leq f(z^k) \leq f(z^0) + 2L\tau m(\epsilon_0 - \epsilon_k)$$

and hence $z^k, \hat{z}^k \in \tilde{\mathcal{F}}_{\epsilon_0}$ and $z^k, \tilde{z}^k \in Z$. ■

5 Some Numerical Experience

To gain some understanding of the practical performance of the ϵ-active set algorithm, we implemented in Matlab a version of this algorithm for the case of quadratic f. We report below the implementation details and our numerical experience.

In our implementation, we set $\epsilon_0 = \nu_0 = 0$ since f is quadratic. To account for roundoff errors, any number below 10^{-12} in magnitude is treated as zero. In Step 1, we set $\hat{z}^k$ to be a stationary point of the quadratic program (3), with $(v_i^k)_{i=1}^m$ and $(w_i^k)_{i=1}^m$ being Lagrange multipliers associated with the first four sets of equations/inequalities. This is implemented by calling the quadratic program solver from the Matlab Optimization Toolbox (Version 1.5.2), with z^k as the starting point. In Step 4, the index i_k is chosen to minimize $\min\{v_i^k, w_i^k\}$ among all indices $i \in A^k \cap B^k$. We choose d^k to minimize $\nabla f(\hat{z}^k)^T d$ among all d satisfying the constraints of (6), ignoring the first inequality. This is implemented by calling the linear program solver from the Matlab Optimization Toolbox. We stop at Step 3 of iteration k when $v_i^k \geq 0, w_i^k \geq 0$ for all $i \in A^k \cap B^k$.

In our testing, we use the Matlab program QPECgen (Version 1.1) of Jiang and Ralph [6] to generate test problems. This program generates random instances of mathematical programs with quadratic objective function and affine variational inequality constraints. We set the QPECgen parameters as in Table 8 of [6]. The problems thus generated are special cases of (1) with

$$f \text{ convex quadratic}, \qquad H_i(x,y) = y_i \quad \forall i, \qquad q = 0,$$

where $x \in \Re^{n-m}$ and $y \in \Re^m$. The dimensions n, m, p for the generated problems are shown in Table 1. The values of `second_deg` and `mono_M`, which are QPECgen parameters that control the degree of degeneracy (failure of strict complementarity) and the monotonicity of $[G_i(x,y)]_{i=1}^m$ in y, are also shown.

To generate the feasible starting point z^0, we considered solving

$$\begin{array}{lll} \text{minimize} & \sum_{i=1}^m G_i(z)H_i(z) & \\ \text{subject to} & G_i(z) \geq 0 & i = 1, \ldots, m, \\ & H_i(z) \geq 0 & i = 1, \ldots, m, \\ & g_j(z) \leq 0 & j = 1, \ldots, p \end{array}$$

for a global optimal solution. However, the Matlab quadratic program solver was unable to find a global optimal solution of this nonconvex quadratic program. Thus, we used a less direct scheme for generating z^0: We first modify the problem by negating the linear terms in the objective function. Then we apply the ϵ-active set algorithm to the modified problem, initialized with the recommended solution that is provided by QPECgen. This generates a feasible solution of (1) which we take to be z^0. [We cannot use the recommended solution as z^0 because it is already a B-stationary point.]

The performance of the algorithm on the test problems is reported in Table 1. As can be seen from Table 1, the algorithm terminated in a finite number of iterations on each problem. Moreover, on all except problems 9 and 10, the final z is verified to be a B-stationary point since the gradients of the active constraints were linearly independent. [In particular, the determinant of $C^T C$, where C is the matrix with columns being the gradients of the active constraints, is a large positive number. On problems 9 and 10, this determinant is below 10^{-12} in magnitude and hence is treated as zero.] Since f is convex, each B-stationary point is a local optimal solution of MPEC (1). For the first 8 problems, the final f-value agrees with those reported in Table 8 of [6] up to the fifth significant digit. The number of iterations increases significantly with n and m, especially on the two largest monotone problems. The number of iterations in which Step 2 is exited is also higher on the large monotone problems. The reasons for this is not well understood, but it does not appear to be related to degeneracy. For example, on the last problem of Table 1, the final solution has 3 degenerate indices, and yet the number of iterations is small relative to the problem size. The work at each iteration k varies, depending on the effort to solve the quadratic program (3) using the warm starting point z^k and the effort

Table 1: Performance of the ϵ-active set algorithm on QPECgen problems.

(n,m,p)	second_deg	mono_M	iter[1]	iter0[2]	initial f[3]	final f[4]	deg[5]
$(58,50,4)$	0	1	15	2	-45.682	-142.823	0
$(108,100,4)$	0	1	17	5	-620.389	-664.385	0
$(158,150,4)$	0	1	40	8	-475.081	-535.741	0
$(208,200,4)$	0	1	278	39	-55.876	-109.594	0
$(58,50,4)$	4	1	16	2	61.830	-41.876	4
$(108,100,4)$	4	1	16	4	-555.222	-599.936	4
$(158,150,4)$	4	1	44	12	-450.023	-536.443	4
$(208,200,4)$	4	1	210	31	19.916	-23.781	4
$(58,50,4)$	4	0	18	3	-78.398	-137.869	1
$(108,100,4)$	4	0	18	1	-165.875	-245.399	0
$(158,150,4)$	4	0	36	2	-349.095	-382.025	1
$(208,200,4)$	4	0	26	1	-39.048	-39.674	3

[1] iter is the number of iterations until termination.
[2] iter0 is the number of iterations in which Step 2 is exited.
[3] This is $f(z^0)$.
[4] This is $f(z^k)$, where k indexes the last iteration.
[5] This is the cardinality of $A^k \cap B^k$, where k indexes the last iteration.

to solve the linear program derived from the system (6). In the case where we exit from Step 2, we can skip solving subproblem (3) at the next iteration $k+1$. This is because $A^{k+1} \supseteq A^k$, $B^{k+1} \supseteq B^k$, so that $\hat{z}^k$ is still a stationary point of (3) when "k" is replaced by "$k+1$."

6 Conclusion

We have proposed an active-set algorithm for solving mathematical programs with linear complementarity constraints and established convergence to a B-stationary point of the problem under the uniform LICQ on the ϵ-feasible set. To the authors' knowledge, this is the first implementable algorithm that enjoys global convergence to a B-stationary point without nondegeneracy (strict complementarity) assumption. We have also reported some numerical results that support the theoretical advantage of the algorithm.

References

[1] Facchinei, F., Jiang, H. and Qi, L., *A smoothing method for mathematical programs with equilibrium constraints*, Math. Programming, 85 (1999), 107–134.

[2] Fukushima, M., Luo, Z.-Q. and Pang, J.-S., *A globally convergent sequential quadratic programming algorithm for mathematical programs with linear complementarity constraints*, Comput. Optim. Appl., 10 (1998), 5–34.

[3] Fukushima, M. and Pang, J.-S., *Convergence of a smoothing continuation method for mathematical programs with complementarity constraints*, in Ill-posed Variational Problems and Regularization Techniques, Lecture Notes in Economics and Mathematical Systems, Vol. 477, M. Thera and R. Tichatschke, eds., Springer-Verlag, Berlin/Heidelberg, 1999.

[4] Hoffman, A. J., *On approximate solutions of systems of linear inequalities*, J. Res. Natl. Bur. Standards, 49 (1952), 263–265.

[5] Jiang, H. and Ralph, D., Smooth SQP methods for mathematical programs with quadratic objectives and affine variational inequality constraints, Manuscript, Department of Mathematics and Statistics, The University of Melbourne, Parkville, AU, 1997.

[6] Jiang, H. and Ralph, D., *QPECgen, a MATLAB generator for mathematical programs with nonlinear complementarity constraints*, Comput. Optim. Appl., 13 (1999), 25–59.

[7] Luo, Z.-Q., Pang, J.-S. and Ralph, D., *Mathematical Programs with Equilibrium Constraints*, Cambridge University Press, Cambridge, 1996.

[8] Luo, Z.-Q. and Tseng, P., *Error bound and convergence analysis of matrix splitting algorithms for the affine variational inequality problem*, SIAM J. Optim., 2 (1992), 43–54.

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

[10] Pang, J.-S. and Fukushima, M., *Complementarity constraint qualifications and simplified B-stationarity conditions for mathematical programs with equilibrium constraints*, Comput. Optim. Appl., 13 (1999), 111–136.

[11] Scholtes, S., Convergence properties of a regularization scheme for mathematical programs with complementarity constraints, Working paper, Judge Institute of Management Studies, University of Cambridge, Cambridge, UK, 1999.

[12] Scholtes, S. and Stöhr, M., *Exact penalization of mathematical programs with equilibrium constraints*, SIAM J. Control Optim., 37 (1999), 617–652.

[13] Scholtes, S. and Stöhr, M., Mathematical programs with complementarity constraints: stationarity, optimality, and sensitivity, Working paper, Judge Institute of Management Studies, University of Cambridge, Cambridge, UK, 1998.