Global convergence of elastic mode approaches for a class of MPCC

Mihai Anitescu
Argonne National Laboratory
Recently, there have been several approaches to solve Mathematical Programs with Complementarity Constraints (MPCC) by using nonlinear programming techniques (General: Anitescu 2000, Fletcher and al 2002; Structured smoothing: Fukushima and Pang 1998, Scholtes 2002).

However all of them are of the local type: If point is sufficiently close to a strongly stationary point that satisfies some condition then algorithm converges to that point.

Global convergence: If algorithm is applied to a problem class then any accumulation point is a stationary point. If the point satisfies some condition then it is a ++ stationary point.

However, we need to restrict the problem class to get some significant results.
Before anything else: The mixed P property

Let $A \in \mathbb{R}^{(n_c+l) \times n_c}$, $B \in \mathbb{R}^{(n_c+l) \times n_c}$, and $C \in \mathbb{R}^{(n_c+l) \times l}$. $[A \ B \ C]$ is mixed P partition if

$$0 \neq (y, w, z) \in \mathbb{R}^{2n_c+l},$$

$$Ay + Bw + Cz = 0$$

$$\Rightarrow \exists i, \ 1 \leq i \leq n_c, \ such \ that \ y_iw_i > 0.$$ 

What is actually needed in this work (and is implied if $[A \ B \ C]$ is a mixed P partition), is

$$A^T \theta \leq 0, \ B^T \theta \leq 0, \ C^T \theta = 0 \ \Rightarrow \ \theta = 0$$
Optimization of mixed P variational inequalities

\[(\text{OMPV})\]
\[
\begin{align*}
\min_{x, y, w, z} & \quad f(x, y, w, z) \\
\text{subj.to} & \quad g(x) \leq 0 \\
& \quad h(x) = 0 \\
& \quad F(x, y, w, z) = 0 \\
& \quad y, w \leq 0 \\
& \quad y^T w \leq 0
\end{align*}
\]

\[(\text{OMPV}(c))\]
\[
\begin{align*}
\min_{x, y, w, z, \zeta_1, \zeta_2} & \quad f(x, y, w, z) + c(\zeta_1 + \zeta_2) \\
\text{subj.to} & \quad g(x) \leq 0 \\
& \quad h(x) = 0 \\
& \quad -\zeta_1 e_{nc+l} \leq F(x, y, w, z) \leq \zeta_1 e_{nc+l} \\
& \quad y, w \leq 0 \\
& \quad y^T w \leq \zeta_2 \\
& \quad \zeta_1, \zeta_2 \geq 0
\end{align*}
\]

We name the problem **OMPV** because of the **mixed P VI**:

\[
F(x, y, w, z) = 0 \quad y, w \leq 0 \quad y^T w \leq 0
\]
MPEC stationarity concepts

\[ \nabla_x f(x, y, w, z)^T + \nabla_x h(x)^T \lambda + \nabla_x g(x)^T \mu + \nabla_x F(x, y, w, z)^T \theta = 0 \]
\[ \nabla_y f(x, y, w, z)^T + \hat{\eta}_y + \nabla_y F(x, y, w, z)^T \theta = 0 \]
\[ \nabla_w f(x, y, w, z)^T + \hat{\eta}_w + \nabla_w F(x, y, w, z)^T \theta = 0 \]
\[ \nabla_z f(x, y, w, z)^T + \nabla_z F(x, y, w, z)^T \theta = 0. \]

\[ g(x) \leq 0, \mu \geq 0, h(x) = 0, g(x)^T \mu = 0 \]

\[ F(x, y, z, w) = 0, y \leq 0, w \leq 0, y^T w = 0, \]

\[ \sum_{k=1}^{n_c} y_k |\hat{\eta}_{y,k}| = 0, \sum_{k=1}^{n_c} w_k |\hat{\eta}_{w,k}| = 0 \]
MPEC stationarity concepts

• Weakly stationary points: no additional requirements.

• C-stationary points: \( \hat{\eta}_{y,k} \hat{\eta}_{w,k} \geq 0, \ k = 1, 2, \ldots, n_c \):

• M-stationary points: C-stationary points and \( \hat{\eta}_{y,k} \geq 0 \) or \( \hat{\eta}_{w,k} \geq 0, \ k = 1, 2, \ldots, n_c \)

• B-stationary points, for which \( d = 0 \) is a solution of the linearized (OMPV) except \( y^T w \leq 0 \)

• Strongly stationary points,

\[ y_k = 0, \ w_k = 0 \Rightarrow \hat{\eta}_{y,k} \geq 0 \text{ and } \hat{\eta}_{w,k} \geq 0, \ k = 1, 2, \ldots, n_c \]

Sheel and Scholtes 2000 describe in detail the connections.
Important concepts about MPCC and OMPV

- **Definition (ULSC)**. A weakly stationary point \((x, y, z, w)\) of (OMPV) satisfies the upper level strict complementarity (ULSC) property if there exists an MPCC multiplier that satisfies

\[
y_k + w_k = 0 \Rightarrow \hat{\eta}_{y,k}\hat{\eta}_{w,k} \neq 0, \quad k = 1, 2, \ldots, n_c.
\]

- **Definition (MPCC-LICQ)** MPCC-LICQ holds at a feasible \((x, y, z, w)\), point of (OMPV) if the gradients of all active constraints of (OMPV) at \((x, y, z, w)\), with the exception of the complementary constraint \(y^Tw \leq 0\), are linearly independent.

**Note (Sheel and Scholtes 2000)** Under MPCC-LICQ, all stationarity concepts are the same at a solution point of (OMPV).
Assumptions

A1 The mappings $f, g, h, F$ are twice continuously differentiable.

A2 The constraints involving only the parameters $x$ satisfy, for any $x$,

i) $\nabla_x h(x)$ has full column rank.

ii) $\exists p \in \mathbb{R}^n$ such that $\nabla_x h(x)p = 0$ and $\nabla g_i(x)p < 0$ whenever $g_i(x) \geq 0$.

iii) The linearization $h(x) + \nabla_x h(x)d = 0$, $g(x) + \nabla_x g(x)d \leq 0$ is feasible.

A3 The partition $[\nabla_y F, \nabla_w F, \nabla_z F]$ is a mixed P partition (3).
Assumptions about the algorithm

Definition (Global Convergence Safeguard). A nonlinear programming algorithm (such as FilterSQP) whose outcome is

1. An infeasible point of the nonlinear program at which the linearization of the constraints is infeasible.

2. A feasible point of the nonlinear program that does not satisfy MFCQ.

3. A feasible point of the nonlinear program that satisfies MFCQ and that is a KKT point of the nonlinear program.

Alg1 The nonlinear programming algorithm has a global convergence safeguard.

Then any accumulation point of a nonlinear programming algorithm that satisfies Assumption Alg1 and is applied to (OMPV(c)) is a KKT point!
$(x, y, w, z, \zeta_1, \zeta_2)$ is an $\varepsilon$ stationary point of (OMPV(c)) if there exists $(\lambda, \mu, \theta, \eta_y, \eta_w, \alpha_c, \alpha_1, \alpha_2)$ such that:

\[
\begin{aligned}
\nabla_x f(x, y, w, z)^T + \nabla_x h(x)^T \lambda + \\
\nabla_x g(x)^T \mu + \nabla_x F(x, y, w, z)^T (\theta^+ - \theta^-) &= t_x \\
\n\nabla_y f(x, y, w, z)^T + \eta_y + \alpha_c w + \nabla_y F(x, y, w, z)^T (\theta^+ - \theta^-) &= t_y \\
\n\nabla_w f(x, y, w, z)^T + \eta_w + \alpha_c y + \nabla_w F(x, y, w, z)^T (\theta^+ - \theta^-) &= t_w \\
\n\nabla_z f(x, y, w, z)^T + \nabla_z F(x, y, w, z)^T (\theta^+ - \theta^-) &= t_z \\
\|\theta^+\|_1 + \|\theta^-\|_1 + \alpha_1 = c + t_{\alpha_1}; \quad \alpha_c + \alpha_2 = c + t_{\alpha_2}
\end{aligned}
\]

$\mu \geq 0; \ \eta_y, \eta_w \geq 0; \ \theta^+, \theta^- \geq 0; \ \alpha_c, \alpha_1, \alpha_2 \geq 0,$

$\|t_x, t_y, t_w, t_z, t_{\alpha_1} t_{\alpha_2}\|_{\infty} \leq \varepsilon.$
\[ \varepsilon \text{ stationary point, primal and compl. conditions} \]

\[
\begin{cases}
  g(x) & \leq t_g \\
  h(x) & = t_h \\
  -\zeta_1 e_{n_c+l} - t_1F & \leq F(x, y, w, z) \leq \zeta_1 e_{n_c+l} + t_2F \\
  y, w & \leq 0 \\
  y^T w & \leq \zeta_2 + t_c \\
  \zeta_1, \zeta_2 & \geq 0,
\end{cases}
\]

\[
\begin{cases}
  (\zeta_1 e_{n_c+l} - F)^T \theta^+ + (F + \zeta_1 e_{n_c+l})^T \theta^- = t_cF \\
  \alpha_c (\zeta_2 - w^T y) = t_{cc}; \quad g(x)^T \mu = t_{cg}; \\
  |\alpha_2 \zeta_2| \leq t_{cp}; \quad |\alpha_1 \zeta_1| \leq t_{cp}; \quad |y^T \eta_y| \leq t_{cp}; \quad |w^T \eta_w| \leq t_{cp},
\end{cases}
\]

\[ \|t_g, t_h, t_1F, t_2F, t_c, t_{cc}, t_{cF}, t_{cg}, t_{cp}\|_{\infty} \leq \varepsilon. \]

... piece of cake for interior-point methods
The algorithm

Choose $c_0 > 0$, $n = 0$, $K > 1$, an integer $q \geq 1$ and a sequence $\varepsilon^n \to 0$.

MPCC Find an $\varepsilon^n$ solution $(x^n, y^n, w^n, z^n, \zeta_1^n, \zeta_2^n)$ of $(\text{OMPV}(c^n))$.

If $\zeta_1^n + \zeta_2^n > (\varepsilon^n)^{\frac{1}{q}}$,

update $c$: $c^{n+1} = Kc^n$ and $n$: $n = n + 1$.

return to MPCC

Note that, as opposed to Scholtes 2002, we do not need an infinite number of steps to solve the subproblem.
Global Convergence Theorem

Assume that

- (OMPV) satisfies the assumptions A1, A2 and A3.
- \((\text{OMPV}(c^n))\) is solved with an NLP algorithm that satisfies Assumption Alg1 that produces an \(\varepsilon^n\) stationary point.
- \(\lim_{n \to \infty} c^n \varepsilon^n = 0\).
- The sequence \((x^{c^n}, y^{c^n}, w^{c^n}, z^{c^n}, \zeta^{c^n}_1, \zeta^{c^n}_2)\) has an accumulation point.

Then (1) if the penalty parameter update rule is activated a finite number of times any accumulation point is a strongly stationary point of (OMPV) and (2) if the penalty parameter update rule is activated an infinite number of times, and then any accumulation point is a \(C\)-stationary point of (OMPV).

Note that we may still diverge to \(\infty\) ... but we’ll fix that.
Approximate second-order stationary points

Definition (ε, χ second-order stationary point). We say that the point \( \tilde{x} = (x, y, z, w, \zeta_1, \zeta_2) \), together with a Lagrange multiplier \( \tilde{\lambda} = (\lambda, \mu, \theta^+, \theta^-, \eta_y, \eta_w, \alpha_c, \alpha_1, \alpha_2) \) is an \( \varepsilon, \chi \) second-order point of (OMPV(\( c \))) if

1. \( \tilde{x} = (x, y, z, w, \zeta_1, \zeta_2) \), is an \( \varepsilon \) stationary point of (OMPV(\( c \))), that satisfies exactly the primal-dual complementarity involving the slack variables \( \eta^T_{y,k} y = 0, \eta^T_{w,k} w = 0 \).

2. \( u^T \Lambda^c_{xx} (\tilde{x}, \tilde{\lambda}) u > 0 \) for any \( u \) that is at the same time in the null space of the gradients of the active bound constraints of (OMPV(\( c \))) and null space of a subset of the \( \chi \)-active non-bound constraints of (OMPV(\( c \))).

Note that sufficient conditions can be tested by by active set methods with rank-revealing factorization.
M-stationarity Result

Assume that

- The problem (OMPV) satisfies assumptions $A1$, $A2$ and $A3$
- $(OMPV(c^n))$ is solved with an algorithm that satisfies Assumption $Alg1$.
- $\tilde{x}^n = (x^n, y^n, z^n, w^n, \zeta_1^n, \zeta_2^n)$ is a $\varepsilon^n, \chi^n$ second-order stationary point of $(OMPV(c^n))$, for all $n = 1, 2, \ldots, \infty$
- $\lim_{n \to \infty} c^n = \infty, \lim_{n \to \infty} \varepsilon^n = 0, \lim_{n \to \infty} \chi^n = 0$ and $\lim_{n \to \infty} c^n \varepsilon^n = 0$.
- $(x^*, y^*, z^*, w^*, \zeta_1^*, \zeta_2^*)$ is an accumulation point of this sequence.
- If $(x^*, y^*, z^*, w^*)$ satisfies MPCC-LICQ,

then $(x^*, y^*, z^*, w^*)$ must be an M-stationary point of (OMPV).
Convergence to strongly stationary points

If, in addition to the assumptions of M-stationarity convergence we have that ULSC holds at the accumulation point \((x^*, y^*, z^*, w^*)\), then \((x^*, y^*, z^*, w^*)\) is a strongly stationary point and, as a result, a strongly stationary point.

The result is similar to the results of Fukushima and Pang 98 and Scholtes 2002, except that it works with approximate points. Sven’s objection However, if ULSC does not hold a descent direction may still exist.
Is M-stationarity sufficient?

Assume that \((x^*, y^*, z^*, w^*)\) is an M-stationary point of (OMPV). Then, for any \(\delta > 0\), the following exist

1. A perturbation \(f^\delta (x, y, w, z)\) of the objective function \(f(x, y, w, z)\) that satisfies \(\| \nabla \tilde{x} f^\delta (x, y, w, z) - \nabla \tilde{x} f(x, y, z, w) \| \leq \delta\) for all \(\tilde{x} = (x, y, z, w)\) in a neighborhood of \((x^*, y^*, z^*, w^*)\).

2. A vector \(l^\delta_F\) that satisfies \(\| l^\delta_F \| \leq \delta\).

3. A point \((x^\delta, y^\delta, z^\delta, w^\delta)\) that satisfies \(\| (x^\delta, y^\delta, z^\delta, w^\delta) - (x^*, y^*, z^*, w^*) \| \leq \delta\) and that is a strongly stationary point (thus a B-stationary point) for the perturbed problem.
The perturbed problem

\[
\begin{align*}
\min_{x,y,w,z} & \quad f^\delta(x, y, w, z) \\
\text{sbj.to} & \quad g(x) \leq 0 \\
& \quad h(x) = 0 \\
& \quad F(x, y, w, z) = l^\delta_F \\
& \quad y, w \leq 0 \\
& \quad (y^T w = 0) \quad y^T w \leq 0
\end{align*}
\]

M-stationary points may be indistinguishable in finite arithmetic or for finite tolerance from strongly-stationary points!
Finishing global convergence: keep iterates finite

A4 The penalty function \( \psi(x, y, w, z) = ||F(x, y, w, z)||_\infty + y^T w \) has bounded level sets over the set defined by the constraints \( g(x) \leq 0, \ h(x) = 0, \ y \leq 0, \ w \leq 0. \)

A5 The objective function \( f(x, y, w, z) \) is bounded below over the same set.

Alg2 For any fixed value of \( c \), the algorithm that is applied for solving the problem (OMPV(\( c \)) decreases the merit function \( f(x, y, z, w) + c\psi(x, y, z, w). \)

The merit function \( \Psi(x, y, w, z, c) = \frac{1}{c} (f(x, y, w, z) - B_f) + \psi(x, y, w, z) \) is always decreasing (even at penalty update) and has bounded level sets \( \Rightarrow \) convergence to C-stationary points is guaranteed!
**The obstacle problem**

\[
\begin{align*}
\min_{x,y,w,z} & \quad f(x, z) \\
\text{sbj.to} & \quad g(x) \leq 0 \\
(\text{OBST}) & \quad -A(x)z + \phi(x) = y \\
& \quad k(\phi(x) - A(x)z) + \chi(x) - z = w \\
& \quad y, w \leq 0 \\
& \quad (y^T w = 0) \quad y^T w \leq 0
\end{align*}
\]

We proved that the obstacle problem satisfies assumptions A1, A2, A3, A4 !!! So not so outlandish after all.
A graph of the obstacle problem

Packaging with rigid parabolic obstacle.
The obstacle problem test set (THANKS SVEN!!)

All of them satisfy Assumption A5

- **The incidence set identification problem** The contact region must be as close as possible to a prescribed shape.

- **The packaging problem with compliant obstacle**. Minimize the area of the membrane, while keeping the membrane in contact with the obstacle over at least a prescribed region.

- **The packaging problem with rigid obstacle**. Same as before but the obstacle is rigid.
Algorithmic choices for our numerical simulations

1. We use **knitro** to solve OMPV(c), the relaxed problem. **knitro** was not proven to satisfy **Alg1**, but we can test for $\epsilon$ stationarity and **knitro** provided one for any problem.

2. $q = 2$, $K = 10$, $c_0 = 10$, and $\epsilon^n = 10^{-3}12^{-n}$. We put $\epsilon^n = \text{opttol} = \text{feastol}$.

3. Stopping Criteria $\zeta_1^n + \zeta_2^n \leq 1e - 7.$

4. **Note that** $c^n \leq 10^{n+1}$, **means that** $c^n \epsilon^n \rightarrow 0$, as required by our results!!
Detecting C-stationarity and M-stationarity

- We construct what we hope are good MPEC multipliers:
  \[
  \hat{\eta}_{w,k} = \eta_{w,k} + cy_k, \quad \hat{\eta}_{y,k} = \eta_{y,k} + cw_k, \quad k = 1, 2, \ldots, n_C.
  \]

- We define
  \[
  \text{Cstat} = \min_{k=1,2,\ldots,n_C} \hat{\eta}_{w,k} \hat{\eta}_{y,k}, \quad \text{Mstat} = \max_{k=1,2,\ldots,n_C} \min \{\hat{\eta}_{w,k}, \hat{\eta}_{y,k}\}.
  \]

- If Cstat \(\geq 0\) we go to a C-stationary point; if Mstat \(\leq 0\), we have also an M-stationary point (Note that the MacMPEC library uses nonnegativity constraints, as opposed to nonpositivity as used here).
**Numerical Results**

<table>
<thead>
<tr>
<th>Problem</th>
<th>Obj</th>
<th>Uc</th>
<th>Ut</th>
<th>Cstat</th>
<th>Mstat</th>
<th>Feval</th>
<th>KFeval</th>
</tr>
</thead>
<tbody>
<tr>
<td>is-1-8</td>
<td>2.352e-08</td>
<td>0</td>
<td>5</td>
<td>4.10e-11</td>
<td>2.89e-09</td>
<td>204</td>
<td>390</td>
</tr>
<tr>
<td>is-1-16</td>
<td>8.639e-06</td>
<td>1</td>
<td>6</td>
<td>9.38e-08</td>
<td>7.85e-06</td>
<td>451</td>
<td>4001</td>
</tr>
<tr>
<td>is-1-32</td>
<td>5.904e-06</td>
<td>2</td>
<td>7</td>
<td>3.36e-08</td>
<td>5.52e-05</td>
<td>2906</td>
<td>1097</td>
</tr>
<tr>
<td>is-2-8</td>
<td>4.517e-03</td>
<td>1</td>
<td>6</td>
<td>5.12e-08</td>
<td>2.84e-07</td>
<td>302</td>
<td>1712</td>
</tr>
<tr>
<td>is-2-16</td>
<td>3.006e-03</td>
<td>1</td>
<td>6</td>
<td>1.27e-06</td>
<td>1.02e-03</td>
<td>434</td>
<td>4001</td>
</tr>
<tr>
<td>is-2-32</td>
<td>1.774e-03</td>
<td>2</td>
<td>5</td>
<td>1.01e-05</td>
<td>3.54e-03</td>
<td>2083</td>
<td>4001</td>
</tr>
<tr>
<td>pc-1-8</td>
<td>6.000e-01</td>
<td>1</td>
<td>5</td>
<td>6.32e-14</td>
<td>1.40e-03</td>
<td>75</td>
<td>4001</td>
</tr>
<tr>
<td>pc-1-16</td>
<td>6.169e-01</td>
<td>1</td>
<td>7</td>
<td>3.82e-21</td>
<td>5.65e-07</td>
<td>297</td>
<td>4001</td>
</tr>
<tr>
<td>pc-1-32</td>
<td>6.529e-01</td>
<td>2</td>
<td>6</td>
<td>9.60e-18</td>
<td>8.93e-05</td>
<td>4999</td>
<td>3081</td>
</tr>
<tr>
<td>pc-2-8</td>
<td>6.731e-01</td>
<td>1</td>
<td>5</td>
<td>1.01e-19</td>
<td>3.03e-06</td>
<td>78</td>
<td>1421</td>
</tr>
<tr>
<td>pc-2-16</td>
<td>7.271e-01</td>
<td>2</td>
<td>5</td>
<td>3.60e-16</td>
<td>1.77e-03</td>
<td>289</td>
<td>1358</td>
</tr>
<tr>
<td>pc-2-32</td>
<td>7.826e-01</td>
<td>2</td>
<td>6</td>
<td>1.84e-16</td>
<td>1.22e-04</td>
<td>645</td>
<td>1350</td>
</tr>
<tr>
<td>pr-1-8</td>
<td>7.879e-01</td>
<td>1</td>
<td>6</td>
<td>9.28e-18</td>
<td>1.03e-06</td>
<td>193</td>
<td>81</td>
</tr>
<tr>
<td>pr-1-16</td>
<td>8.260e-01</td>
<td>2</td>
<td>5</td>
<td>1.68e-16</td>
<td>1.14e-05</td>
<td>218</td>
<td>54</td>
</tr>
<tr>
<td>pr-1-32</td>
<td>8.508e-01</td>
<td>2</td>
<td>5</td>
<td>1.95e-17</td>
<td>1.17e-03</td>
<td>644</td>
<td>3040</td>
</tr>
<tr>
<td>pr-2-8</td>
<td>7.804e-01</td>
<td>1</td>
<td>6</td>
<td>3.20e-18</td>
<td>1.46e-06</td>
<td>183</td>
<td>33</td>
</tr>
<tr>
<td>pr-2-16</td>
<td>1.085e+00</td>
<td>3</td>
<td>6</td>
<td>2.32e-15</td>
<td>1.73e-05</td>
<td>342</td>
<td>208</td>
</tr>
<tr>
<td>pr-2-32</td>
<td>1.135e+00</td>
<td>3</td>
<td>6</td>
<td>1.36e-14</td>
<td>1.59e-04</td>
<td>661</td>
<td>2792</td>
</tr>
</tbody>
</table>

**Note** C-stationarity always satisfied, M-stationarity almost true.
M-stationary points under finite tolerance

- The problem pr-1-32, for index $k = 19$ we have
  $$y_{19} = 1.039e-05, \ w_{19} = 1.42e-04, \ \hat{\eta}_{y,19} = 0.14, \ \hat{\eta}_{w,19} = 1.17e-03$$

- In absence of any additional information (such as whether MPCC-LICQ holds, which cannot be tested for AMPL), it is difficult to decide whether the algorithm converges to an M-stationary point at which descent is still possible, or whether it converges to a strongly stationary point.

- However, if MPCC-LICQ holds, then one should somehow take advantage of Sven’s point. But how to do that before convergence, is not clear.
The performance plot

% of problems solved no more than x times slower than the best solver

% of problems solved

percentage of problems

x

X

0 10 20 30 40 50 60

0 10 20 30 40 50 60

30 40 50 60 70 80 90 100

elastic knitro

knitro

27
Conclusions

• We proved that an elastic mode approach are guaranteed to converge to C-stationary points of the optimization of mixed P variational inequalities. To my knowledge, the first that does not assume any other constraint qualification at the solution.

• We proved that several variants of the obstacle problem satisfy our convergence assumptions.

• We have shown that M-stationary points can be confounded with strongly stationary points in finite arithmetic. This does not mean that they will be but in some of our examples they were.

• We have shown that our elastic mode approach with knitro solving the relaxed problem is superior to knitro alone at solving the problem.
Still to do

- Can one robustly marry this approach with an active-set approach to take advantage of MPCC-LICQ (if it holds) sufficiently close to the solution?
- Can convergence to M-stationarity hold under weaker conditions? For example MPCC-MFCQ (see Jane Ye’s talk from Sunday)?