

On ϵ -regularized solutions for MPCC.

Jean Pierre Dussault Mounir Haddou Abdeslam Kadrani
Tangi Migot

November 22, 2016

Abstract

We discuss requirements for approximate solutions of regularized sequences of MPCC to converge to M-stationary points. To compute those approximate solutions, we present an hybrid approach using penalization and active sets quite general, encompassing KDB, KS and the butterfly regularizations.

Keywords: Complementarity constraints, MPCC, regularization, nonlinear Programming

We study the numerical solution of the mathematical programs with complementarity constraints (MPCC)

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g(x) \leq 0, \\ & h(x) = 0, \\ & 0 \leq G(x) \perp H(x) \geq 0, \end{aligned} \tag{1}$$

where $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $g : \mathbb{R}^n \rightarrow \mathbb{R}^{n_i}$, $h : \mathbb{R}^n \rightarrow \mathbb{R}^{n_e}$ and $G, H : \mathbb{R}^n \rightarrow \mathbb{R}^{n_c}$ are continuously differentiable. The relation $G(x) \perp H(x)$ may be represented by the Hadamar product $G(x) \circ H(x) \stackrel{\text{def}}{=} (G_1 H_1, \dots, G_{n_c} H_{n_c})$.

Using Lagrange multipliers or dual variables ν_g , ν_h , μ_G , μ_H and η , we define the Lagrangian

$$\mathcal{L}(x, \nu_g, \nu_h, \mu_G, \mu_H, \eta) = f(x) + \nu_g g(x) + \nu_h h(x) + \mu_G G(x) + \mu_H H(x) + \eta(\dots). \tag{2}$$

Note that we use the convention that primal quantities are column vectors while their dual counterparts are row vectors.

In general, there does not exist KKT stationary points for (1) since it is highly degenerate and does not satisfy a constraint qualification. A KKT stationary point x^* would imply the existence of $\nu_g^* \leq 0$, ν_h^* , $\mu_G^* \geq 0$, $\mu_H^* \geq 0$, $\eta^* \geq 0$ such that $g(x^*) \leq 0$, $h(x^*) = 0$, $0 \leq G(x^*) \perp H(x^*) \geq 0$, and $\nu_g^* g(x^*) = 0$, $\mu_G^* G(x^*) = 0$, $\mu_H^* H(x^*) = 0$ and $\nabla_x(\mathcal{L}(x^*, \nu_g^*, \nu_h^*, \mu_G^*, \mu_H^*, \eta^*)) = 0$. Such a solution is known as a strong stationary point, S-stationary for short. The

best one can ensure for a local minimum is an M-stationary point, in which the condition $\mu_G \geq 0, \mu_H \geq 0$ must be weakened to

$$\mu^* \geq 0 \text{ and for all } l \in \{i : G_i(x^*) = H_i(x^*) = 0\}, \text{ either } \mu_{G_l}^*, \mu_{H_l}^* > 0, \text{ or } \mu_{G_l}^* \mu_{H_l}^* = 0, \quad (3)$$

Therefore, devising algorithms to reach KKT stationary points (S-stationary) is not possible in general, and we must satisfy ourselves in devising algorithms reaching M-stationary points. The following example due to Kanzow and Schwartz exhibits a situation where the global minimizer is not a KKT point but an M-stationary point. We will return to this example later on.

Example 1

$$\begin{aligned} \min \quad & x_1 + x_2 - x_3 \\ \text{s.t.} \quad & -4x_1 + x_3 \leq 0 \\ & -4x_2 + x_3 \leq 0 \\ & 0 \leq x_1 \perp x_2 \geq 0. \end{aligned} \quad (4)$$

The global solution is $(0, 0, 0)^t$ but is not a KKT point. Indeed, the gradient of the Lagrangian turns out to be

$$\begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} + \nu_1 \begin{pmatrix} -4 \\ 0 \\ 1 \end{pmatrix} + \nu_2 \begin{pmatrix} 0 \\ -4 \\ 1 \end{pmatrix} - \mu_1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} - \mu_2 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + \eta \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \quad (5)$$

and since $\nu_1 + \nu_2 = 1$ (third line), summing the first two lines yields $2 - 4(\nu_1 + \nu_2) - \mu_1 - \mu_2 = 0$ and therefore $\mu_1 + \mu_2 = -2$; both cannot be non-negative.

1 Regularized subproblems

We intend replacing the complementarity constraint $G(x) \circ H(x) \leq 0$ by a suitable regularization/approximation parametrized using t , $\Phi(G, H, t) \leq 0$. The first such regularization was proposed by Scholtes, $\Phi(G, H, t) = G(x) \circ H(x) - te_{n_c}$. Many other such smoothing domains were proposed, all suffering from the weakness that a sequence of regularized stationary points can converge to a C-stationary solution for the original problem. The first approximation yielding convergence to M-stationary points was proposed by Kadrani, Dussault and Benchakroun: $\Phi(G, H, t) = (G(x) - t) \circ (H(x) - t)$. This was followed by a true regularization proposed by Kanzow and Schwartz, and the so called butterfly regularization by Dussault, Haddou and Migot using a further parameter r and taking the form $\Phi(G, H, t, r) = (H(x) - t\theta_r(G(x))) \circ (G(x) - t\theta_r(H(x)))$. Moreover, the $G(x) \geq 0$ and $H(x) \geq 0$ constraints are often relaxed so that the resulting problem takes the form

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g(x) \leq 0, \quad h(x) = 0, \\ & G(x) \geq -t, \quad H(x) \geq -t, \\ & \Phi(G, H, t, r) \leq 0. \end{aligned} \quad (6)$$

We may now characterize KKT points for (6):

$$\begin{aligned}
& \nabla_x \mathcal{L}(x, \nu_g, \nu_h, \mu_G, \mu_H, \eta) = 0, \\
& 0 \leq \nu_g, \quad g(x) \leq 0, \quad \nu_g g(x) = 0, \\
& h(x) = 0, \\
& 0 \leq \mu_G, \quad (G(x) + t \cdot e_c) \geq 0, \quad \mu_G(G(x) + t \cdot e_c) = 0, \\
& 0 \leq \mu_H, \quad (H(x) + t \cdot e_c) \geq 0, \quad \mu_H(H(x) + t \cdot e_c) = 0, \\
& 0 \leq \eta, \quad \Phi(G, H, t, r) \leq 0, \quad \eta \Phi(G, H, t, r) \leq 0.
\end{aligned} \tag{7}$$

1.1 ϵ -solution to regularized subproblems

We now consider approximate KKT points for (6). It consists in replacing most “0” in (7) by small quantities ϵ . However, it is important that the relations on Φ be satisfied exactly, not within ϵ as we discuss below. Moreover, we chose to keep the sign constraints on the various multipliers. The remaining “0” are replaced by ϵ . Each relation could have a dedicated ϵ , but we keep the notation simpler by using a single ϵ . Also, relaxed constraints $G(x) \geq -t$ and $H(x) \geq -t$ are not further relaxed.

Definition 1 *We say that $x \in \mathbb{R}^{n+3n_c+n_i+n_e}$ is an ϵ -stationary point of (1) if there exist multipliers ν, μ, η satisfying*

$$\begin{aligned}
& \|\nabla_x \mathcal{L}(x, \nu_g, \nu_h, \mu_G, \mu_H, \eta)\|_\infty \leq \epsilon, \\
& 0 \leq \nu_g, \quad g(x) \leq \epsilon, \quad \nu_g g(x) \leq \epsilon, \\
& \|h(x)\| \leq \epsilon, \quad |\nu_h h(x)| \leq \epsilon, \\
& 0 \leq \mu_G, \quad (G(x) + t \cdot e_c) \geq 0, \quad \mu_G(G(x) + t \cdot e_c) \leq \epsilon, \\
& 0 \leq \mu_H, \quad (H(x) + t \cdot e_c) \geq 0, \quad \mu_H(H(x) + t \cdot e_c) \leq \epsilon, \\
& 0 \leq \eta, \quad \Phi(G, H, t, r) \leq 0, \quad \eta \Phi(G, H, t, r) \leq 0.
\end{aligned} \tag{8}$$

As noted above, the formulation avoids relaxing the $\Phi(G, H, t, r)$ relations. If the complementarity condition on the complementarity constraint is relaxed to

$$\eta [(G(x) - t \cdot e_c) \circ (H(x) - t \cdot e_c)] \leq \epsilon,$$

Kanzow and Schwartz provide the following example exhibiting convergence to a W-stationary point.

Example 2

$$\begin{aligned}
& \min \quad x_2 - x_1 \\
& \text{s.t.} \quad 0 \leq x_1 \perp x_2 \geq 0.
\end{aligned} \tag{9}$$

If we perturb the relation $\eta \Phi(x_1, x_2, t, r) \leq \epsilon$ (leaving the other conditions $x \geq 0$, $\Phi(x_1, x_2, t, r) \leq 0$), η may be positive when the constraint $\Phi(x_1, x_2, t, r)$ is not active. For the case KDB $\Phi(x_1, x_2, t, r) = (x_1 - t)(x_2 - t) = -\epsilon^2$ with $\epsilon = t^2$, the point $x(t, \epsilon) = \begin{pmatrix} t-\epsilon \\ t+\epsilon \end{pmatrix} \geq \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ is ϵ -stationary for small enough ϵ : $\Phi(x_1, x_2, t, r) \leq \epsilon$ and the choice $\eta = \frac{1}{\epsilon}$ makes the Lagrangian $\begin{pmatrix} -1 \\ 1 \end{pmatrix} + \eta \Phi(x_1, x_2, t, r)$ vanish. $x(t)$ converges to the origin when $t, \epsilon \rightarrow 0$ but the origin is only weakly stationary.

Now, if the complementarity constraint is relaxed to

$$(G(x) - t \cdot e_c) \circ (H(x) - t \cdot e_c) \leq \epsilon,$$

convergence may occur to C-stationary points as shown in the following example.

Example 3

$$\begin{aligned} \min \quad & \frac{1}{2}((x_1 - 1)^2 + (x_2 - 1)^2) \\ \text{s.t.} \quad & 0 \leq x_1 \perp x_2 \geq 0. \end{aligned} \tag{10}$$

We specialize the relations (8) as the following, for t and ϵ close to 0.

$$\begin{aligned} & \left\| \begin{pmatrix} x_1-1 \\ x_2-1 \end{pmatrix} - \mu_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} - \mu_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \eta \begin{pmatrix} x_2-t \\ x_1-t \end{pmatrix} \right\|_{\infty} \leq \epsilon, \\ & 0 \leq \mu_1, \quad (x_1 + t) \geq 0, \quad \mu_1(x_1 + t) \leq \epsilon, \\ & 0 \leq \mu_2, \quad (x_2 + t) \geq 0, \quad \mu_2(x_2 + t) \leq \epsilon, \\ & 0 \leq \eta, \quad (x_1 - t)(x_2 - t) \leq \epsilon, \quad \eta[(x_1 - t)(x_2 - t) - \epsilon] \geq 0. \end{aligned} \tag{11}$$

The points $\begin{pmatrix} t+\sqrt{\epsilon} \\ t+\sqrt{\epsilon} \end{pmatrix}$ together with $\mu_1 = \mu_2 = 0$ and $\eta = \frac{1-t-\sqrt{\epsilon}}{\sqrt{\epsilon}} \nearrow +\infty$ satisfy the above relations. The limit point when $t, \epsilon \rightarrow 0$ is the origin, which is a C-stationary point with $\mu_1 = \mu_2 = -1$.

On this example, the relaxed regularized complementarity constraint is active for any small enough $t, \epsilon > 0$; moreover, the relaxed regularized stationary point is a local maximum for $t + 2\sqrt{\epsilon} < 1$. The origin is a local maximum for the original MPCC

Another example might help understanding the phenomenon.

Example 4

$$\begin{aligned} \min \quad & -\frac{1}{2}((x_1 - 1)^2 + (x_2 - 1)^2) \\ \text{s.t.} \quad & 0 \leq x_1 \perp x_2 \geq 0. \end{aligned} \tag{12}$$

We again specialize the relations (8) as the following, for t and ϵ close to 0.

$$\begin{aligned} & \left\| \begin{pmatrix} 1-x_1 \\ 1-x_2 \end{pmatrix} - \mu_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} - \mu_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \eta \begin{pmatrix} x_2-t \\ x_1-t \end{pmatrix} \right\|_{\infty} \leq \epsilon, \\ & 0 \leq \mu_1, \quad (x_1 + t) \geq 0, \quad \mu_1(x_1 + t) \leq \epsilon, \\ & 0 \leq \mu_2, \quad (x_2 + t) \geq 0, \quad \mu_2(x_2 + t) \leq \epsilon, \\ & 0 \leq \eta, \quad (x_1 - t)(x_2 - t) \leq \epsilon, \quad \eta[(x_1 - t)(x_2 - t) - \epsilon] \geq 0. \end{aligned} \tag{13}$$

This time, the points $\begin{pmatrix} t+\sqrt{\epsilon} \\ t+\sqrt{\epsilon} \end{pmatrix}$ are no more ϵ -stationary but the points $x = \begin{pmatrix} 1 \\ -t \end{pmatrix}$, $\mu_2 = 1 + t$ and $x = \begin{pmatrix} -t \\ 1 \end{pmatrix}$, $\mu_1 = 1 + t$ are. Their limits are $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ which are KKT points for the original MPCC with $\mu = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ or $\mu = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. $\begin{pmatrix} -t \\ -t \end{pmatrix}$ with $\mu = \begin{pmatrix} 1+t \\ 1+t \end{pmatrix}$ is also stationary, and of course converges to the origin, a local minimizer of the original MPCC.

In this example, the limit points are not minimizers for the original MPCC, but satisfy the first order KKT conditions for a minimizer. The second order conditions fails for those limit points. The two examples show limiting solutions of regularized subproblems which are not local minimizers of the original MPCC. The first one fails to satisfy a first order condition while the second one satisfies such a first order condition but not the second order one (it is a maximum on the active set).

1.2 Lagrange multipliers of regularizations

The following example develops on example 1 due to Kanzow and Schwartz and exhibits a situation where the regularized subproblems have no KKT point.

Example 5 *The KDB regularized problem is*

$$\begin{aligned}
\min \quad & x_1 + x_2 - x_3 \\
\text{s.t.} \quad & -4x_1 + x_3 \leq 0 \\
& -4x_2 + x_3 \leq 0 \\
& x_1 \geq -t \\
& x_2 \geq -t \\
& (x_1 - t)(x_2 - t) \leq 0.
\end{aligned} \tag{14}$$

The point $(t, t, 4t)^t$ is feasible so that the minimum value of this program is $\leq -2t$. Moreover, whenever $x_1 > t$, we must have $x_2 \leq t$ to satisfy $(x_1 - t)(x_2 - t) \leq 0$. This allows to conclude that $(t, t, 4t)^t$ is the global minimum of the regularized problem. $\mu = 0$ and the gradient of the Lagrangian yields

$$\begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} + \nu_1 \begin{pmatrix} -4 \\ 0 \\ 1 \end{pmatrix} + \nu_2 \begin{pmatrix} 0 \\ -4 \\ 1 \end{pmatrix} + \eta \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \tag{15}$$

which cannot be satisfied.

This last example seems to contradict theorem 4.6 in [2], but MPCC-LICQ is not satisfied by four constraints in \mathbb{R}^3 .

2 Convergence of ϵ -stationary points

We now address the convergence of sequences of ϵ -stationary points. With the conditions developed in the definition 1, section 1, we are able to prove convergence to M-stationary points.

3 How to compute ϵ -stationary points

We now specialize the MPCC models using slack variables to state complementarity constraints using only variable components. Moreover, we treat all other constraints using penalization.

By introducing slack variables, MPCC (1) can be equivalently written in the form

$$\begin{aligned}
\min \quad & f(x) \\
\text{s.t.} \quad & g(x) + s = 0, \quad s \geq 0, \\
& h(x) = 0, \\
& G(x) - y_G = 0, \quad H(x) - y_H = 0, \\
& y_G \geq 0, \quad y_H \geq 0, \\
& y_G \circ y_H \leq 0.
\end{aligned} \tag{16}$$

This reformulated MPCC has the same properties as (1).

We now penalize all of the equality constraints.

$$\begin{aligned}
\min \quad & f(x) + \frac{1}{2\rho}\phi(x, y, s) \\
\text{s.t.} \quad & s \geq 0, \\
& y \geq 0, \\
& y_G \circ y_H \leq 0,
\end{aligned} \tag{17}$$

where ϕ is the penalty function

$$\phi(x, y, s) = \|(g(x) + s, h(x), G(x) - y_G, H(x) - y_H)\|^2.$$

We now regularize the complementarity constraints to get

$$\begin{aligned}
\min \quad & f(x) + \frac{1}{2\rho}\phi(x, y, s) \\
\text{s.t.} \quad & s \geq 0, \\
& y_G \geq -te, \quad y_H \geq -te, \\
& \Phi(y_G, y_H, t, r) \leq 0.
\end{aligned} \tag{18}$$

Recall that the general form is simplified to $\Phi(y_G, y_H, t, r) = (y_H - t\theta_r(y_G)) \circ (y_G - t\theta_r(y_H))$. This particular form allows to construct an active set strategy involving the “pseudo constraints” $(y_H - t\theta_r(y_G)) = 0$ and $(y_G - t\theta_r(y_H)) = 0$. One of those must be non negative and the other non positive, component by component. When one vanishes, it becomes “active”. The other bounds $y_G \geq -te$ and $y_H \geq -te$ are mutually exclusive with the pseudo constraints, the Φ constraints acting as upper bounds.

When such a bound becomes active, say $y_{H_i} = t\theta_r(y_{G_i})$, its value $(t\theta_r(y_{G_i}))$ may be substituted for y_{H_i} in (18). Similarly, any active s_j is set to 0 and any active $y_{G_i} = -t$ is treated as fixed to $-t$. For such an active set candidate (working set), the essentially unconstrained resulting problem is addressed by any descent method (we recommend ARC). As with any active set method, when a suitable approximate solution is reached on a given working set, a constraint with an estimated Lagrange multiplier of wrong sign is selected to be removed from the working set. In our particular setting, it means freeing a component. When the unconstrained descent method hits an inactive constraint, it is added to the working set.

Therefore, the penalization approach takes care of the various ϵ in the approximate solutions (by reducing ρ for fixed t) while the active set approach ensures the exact satisfaction of the relations involving Φ . The remainder of this section develops the details of this combined penalization–active set approach. Notice that the resulting algorithm generalizes the one in [1] to address non linear θ functions by explicitly resorting to substitution when activating a constraint.

3.1 Active set solution of (18) for fixed ρ

3.2 Strategies to drive ρ and t to zero

References

- [1] Abdeslam Kadrani, Jean Pierre Dussault, and Abdelhamid Benchakroun. A globally convergent algorithm for MPCC. *EURO Journal on Computational Optimization*, 3(3):263–296, 2015.
- [2] Christian Kanzow and Alexandra Schwartz. A new regularization method for mathematical programs with complementarity constraints with strong convergence properties. *SIAM Journal on Optimization*, 23(2):770–798, 2013.