Technische Universität Braunschweig

Mathematical Programs with Complementarity Constraints
Part 2: Relaxation and smoothing-based algorithms

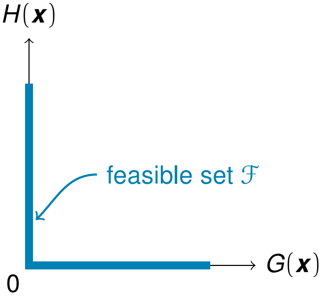
Christian Kirches | Freiburg | September 15, 2023

Overview

- NCP Functions and Subdifferentials
- Relaxations:
  Scholtes, Lin-Fukushima
- Smoothing-Relaxations:
  Steffensen-Ulbrich, Hoheisel
- Kinked Relaxations:
  Kadrani, Kanzow-Schwartz
- Inexactness Effects
- Interior Point Methods
- Sequential Quadratic Programming
- Augmented Lagrangian Methods

Problem Class

$$
\begin{aligned}
\min_{\boldsymbol{x}\in\mathbb{R}^n}\ & F(\boldsymbol{x}) \\
\text{s.t. } & C(\boldsymbol{x}) = \mathbf{0} \\
& D(\boldsymbol{x}) \geqslant \mathbf{0} \\
& \mathbf{0} \leqslant G(\boldsymbol{x}) \perp H(\boldsymbol{x}) \geqslant \mathbf{0}
\end{aligned}
$$



Continuously differentiable $F : \mathbb{R}^n \to \mathbb{R}$, $G, H : \mathbb{R}^n \to \mathbb{R}^c$

Writing "$\mathbf{0} \leqslant \boldsymbol{u} \perp \boldsymbol{v} \geqslant \mathbf{0}$" means to ask that

for all $1 \leqslant i \leqslant c$: $0 = u_i$ **OR** $0 = v_i$ holds.

Vertical Form

Any MPCC can be cast in an **vertical form** that has **orthogonal** complementarities only:

$$\boxed{\begin{aligned} \min_{(\boldsymbol{x},\boldsymbol{u},\boldsymbol{v})\in\mathbb{R}^{n+2c}} \quad & F(\boldsymbol{x}) \\ \text{s.t.} \quad & G(\boldsymbol{x}) - \boldsymbol{u} = \boldsymbol{0} \\ & H(\boldsymbol{x}) - \boldsymbol{v} = \boldsymbol{0} \\ & \boldsymbol{0} \leqslant \boldsymbol{u} \ \perp \ \boldsymbol{v} \geqslant \boldsymbol{0} \end{aligned}}$$

Equivalent Formulations

Under the bounds $\boldsymbol{u} \geqslant 0$, $\boldsymbol{v} \geqslant 0$, several equivalent formulations exist:

- $\boldsymbol{u}^T \boldsymbol{v} = 0$
- $\boldsymbol{u}^T \boldsymbol{v} \leqslant 0$
- $\boldsymbol{u} \circ \boldsymbol{v} = \mathbf{0}$ (Hadamard product)
- $\boldsymbol{u} \circ \boldsymbol{v} \leqslant \mathbf{0}$
- $u_i \cdot v_i = 0$ for all $1 \leqslant i \leqslant c$
- $u_i \cdot v_i \leqslant 0$ for all $1 \leqslant i \leqslant c$

Non-smooth minimization

An **NCP** function $\phi : \mathbb{R}^2 \to \mathbb{R}$ satisfies

$$\phi(u, v) = 0 \iff 0 \leqslant u \perp v \geqslant 0.$$

Using an NCP function, one solves

$$\boxed{\begin{aligned} &\min_{\boldsymbol{x} \in \mathbb{R}^n} F(\boldsymbol{x}) \\ &\text{ s.t. } \phi(G_i(\boldsymbol{x}), H_i(\boldsymbol{x})) = 0,\ 1 \leqslant i \leqslant c \end{aligned}}$$

Useful NCP-functions are **nondifferentiable** in $(0, 0)$.
Differentiable NCP-functions necessarily satisfy $\nabla\phi(0, 0) = (0, 0)^T$.

Bouligand Subdifferential

Denote by D_ϕ the set

$$D_\phi := \{\boldsymbol{x} \mid \phi \text{ is differentiable in } \boldsymbol{x}\}.$$

The set

$$\partial^{\mathrm{B}}\phi(\bar{\boldsymbol{x}}) = \left\{\boldsymbol{d} \;\middle|\; \exists\{\boldsymbol{x}_k\} \subseteq D_\phi,\ \lim_{k\to\infty} \boldsymbol{x}_k = \bar{\boldsymbol{x}} : \lim_{k\to\infty} \phi(\boldsymbol{x}_k) = \boldsymbol{d}\right\}$$

is called the **Bouligand Subdifferential** of ϕ at $\bar{\boldsymbol{x}}$.

Bouligand Subdifferential

Denote by D_ϕ the set

$$D_\phi := \{\boldsymbol{x} \mid \phi \text{ is differentiable in } \boldsymbol{x}\}.$$

The set

$$\partial^{\mathrm{B}}\phi(\bar{\boldsymbol{x}}) = \left\{\boldsymbol{d} \;\middle|\; \exists\{\boldsymbol{x}_k\} \subseteq D_\phi,\ \lim_{k\to\infty} \boldsymbol{x}_k = \bar{\boldsymbol{x}} : \lim_{k\to\infty} \phi(\boldsymbol{x}_k) = \boldsymbol{d}\right\}$$

is called the **Bouligand Subdifferential** of ϕ at $\bar{\boldsymbol{x}}$.

For MPCC with the NCP function $\phi_i(\bar{\boldsymbol{x}}) := \phi(G_i(\bar{\boldsymbol{x}}), H_i(\bar{\boldsymbol{x}}))$ we find:

- $i \in I_{0+}(\bar{\boldsymbol{x}})$: $\partial^{\mathrm{B}}\phi_i(\bar{\boldsymbol{x}}) = \left\{(\nabla G_i(\bar{\boldsymbol{x}}), 0)^T\right\}$
- $i \in I_{+0}(\bar{\boldsymbol{x}})$: $\partial^{\mathrm{B}}\phi_i(\bar{\boldsymbol{x}}) = \left\{(0, \nabla H_i(\bar{\boldsymbol{x}}))^T\right\}$
- $i \in I_{00}(\bar{\boldsymbol{x}})$: $\partial^{\mathrm{B}}\phi_i(\bar{\boldsymbol{x}}) = \left\{(\nabla G_i(\bar{\boldsymbol{x}}), 0)^T, (0, \nabla H_i(\bar{\boldsymbol{x}}))^T\right\}$

Clarke Subdifferential

The set

$$\partial^{C}\phi(\bar{\boldsymbol{x}}) := \operatorname{conv} \partial^{B}\phi(\bar{\boldsymbol{x}})$$

is called the **Clarke Subdifferential** of ϕ at $\bar{\boldsymbol{x}}$. For MPCC with the NCP function $\phi_i(\bar{\boldsymbol{x}}) := \phi(G_i(\bar{\boldsymbol{x}}), H_i(\bar{\boldsymbol{x}}))$ we find:

- $i \in I_{0+}(\bar{\boldsymbol{x}})$: $\partial^{C}\phi_i(\bar{\boldsymbol{x}}) = \partial^{B}\phi_i(\bar{\boldsymbol{x}})$
- $i \in I_{+0}(\bar{\boldsymbol{x}})$: $\partial^{C}\phi_i(\bar{\boldsymbol{x}}) = \partial^{B}\phi_i(\bar{\boldsymbol{x}})$
- $i \in I_{00}(\bar{\boldsymbol{x}})$: $\partial^{C}\phi_i(\bar{\boldsymbol{x}}) = \operatorname{conv}\left\{(\nabla G_i(\bar{\boldsymbol{x}}), 0)^T, (0, \nabla H_i(\bar{\boldsymbol{x}}))^T\right\}$

Chain Rule for ∂^{C} :

$$\partial^{C}(F_1 \circ F_2)(\bar{\boldsymbol{x}}) \cdot \boldsymbol{d} \subseteq \operatorname{conv}(\partial^{C}F_1(F_2(\bar{\boldsymbol{x}})) \cdot \partial^{C}F_2(\bar{\boldsymbol{x}})) \cdot \boldsymbol{d}$$

and equality holds if either F_1 is $\mathcal{C}^1$ around $F_2(\bar{\boldsymbol{x}})$ or F_2 is $\mathcal{C}^1$ around $\bar{\boldsymbol{x}}$.

Scholtes' Relaxation

Solve a sequence of parameterized NLPs for $t \geqslant 0$:

$$(\mathrm{NLP}(t)) \qquad \boxed{\begin{aligned} \min_{\boldsymbol{x} \in \mathbb{R}^n} \ & F(\boldsymbol{x}) \\ \text{s.t. } & C(\boldsymbol{x}) = \mathbf{0} \\ & D(\boldsymbol{x}) \geqslant \mathbf{0} \\ & G_i(\boldsymbol{x}) \cdot H_i(\boldsymbol{x}) \leqslant t, \ 1 \leqslant i \leqslant c \end{aligned}}$$

Theorem

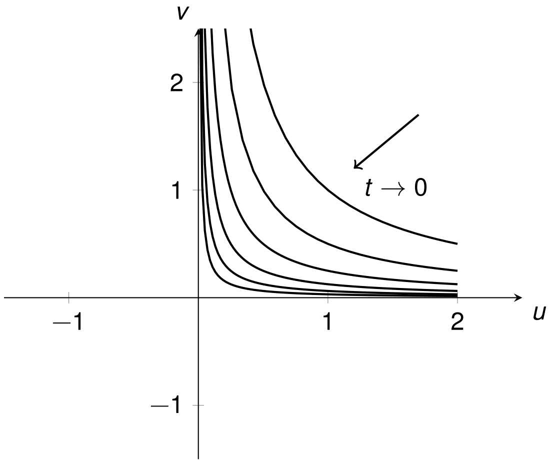
Let $\bar{\boldsymbol{x}} \in \mathbb{R}^n$ be feasible for (MPCC) and let MPCC-MFCQ hold at $\bar{\boldsymbol{x}}$. Then there is an open neighborhood $U(\bar{\boldsymbol{x}})$ and threshold $\bar{t} > 0$ such that for all $t \in [0, \bar{t}]$ one has: If $\boldsymbol{x} \in U(\bar{\boldsymbol{x}})$ is feasible for $\mathrm{NLP}(t)$, then MFCQ holds at $\boldsymbol{x}$.

Theorem

Let $\lim_{k\to\infty} \{t^{(k)}\} = 0$, let $\boldsymbol{x}^{(k)}$ be KKT points of $\mathrm{NLP}(t^{(k)})$ with $\lim_{k\to\infty} \{\boldsymbol{x}^{(k)}\} = \boldsymbol{x}^*$, and let MPCC-MFCQ hold at $\boldsymbol{x}^*$. Then $\boldsymbol{x}^*$ is a C-stationary point.

Under MPCC-LICQ, convergence can also be shown for the unique sequence of MPCC-multipliers.

Scholtes' Relaxation



Scholtes' relaxation for the MPCC constraint $0 \leqslant u \perp v \geqslant 0$.

Scholtes' Relaxation

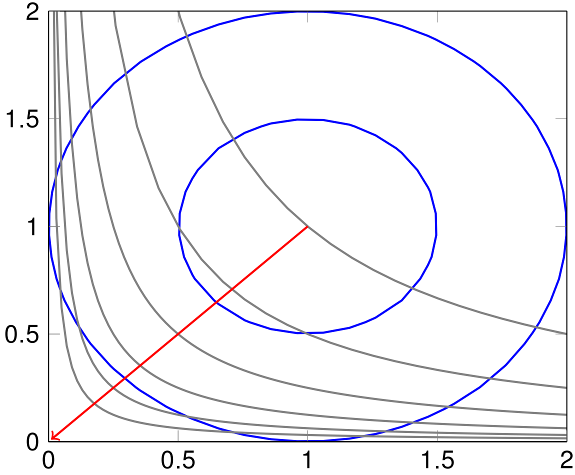
C-stationary is necessary under MPCC-MFCQ. But even assuming MPCC-LICQ does not help. The result is sharp in the following sense:

$$\min_{\boldsymbol{x}\in\mathbb{R}^2} \ (x_1-1)^2+(x_2-1)^2$$
$$\text{s.t. } 0 \leqslant x_1 \ \perp \ x_2 \geqslant 0$$

has two S-stationary local minima at $(0,1)^T$ and $(1,0)^T$, where MPCC-LICQ holds. The local maximum $(0,0)^T$ is C-stationary.

For $t > 0$ sufficiently small, the points $\boldsymbol{x}(t) = (\sqrt{t}, \sqrt{t})^T$ are classical KKT points of the smooth relaxed problem NLP(t).

Scholtes' Relaxation



C-stationarity example for Scholtes' relaxation.

Lin-Fukushima Relaxation

Solve a sequence of parameterized NLPs for $t \geqslant 0$:

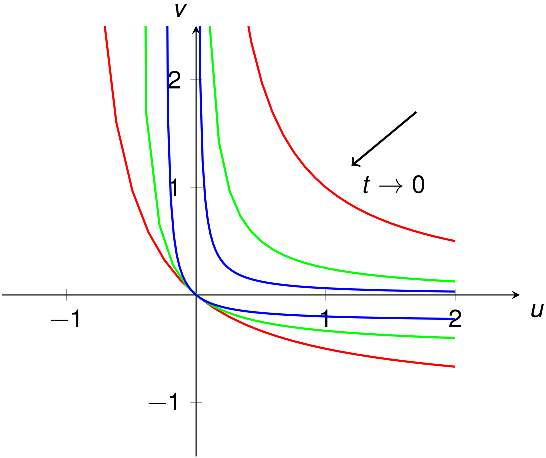
(NLP(t))

$$\begin{aligned} \min_{\boldsymbol{x} \in \mathbb{R}^n} \; & F(\boldsymbol{x}) \\ \text{s.t.} \; & C(\boldsymbol{x}) = \mathbf{0}, \; D(\boldsymbol{x}) \geqslant \mathbf{0} \\ & G_i(\boldsymbol{x}) \cdot H_i(\boldsymbol{x}) \leqslant t^2, \; 1 \leqslant i \leqslant c \\ & (G_i(\boldsymbol{x}) + t) \cdot (H_i(\boldsymbol{x}) + t)) \geqslant t^2, \; 1 \leqslant i \leqslant c \end{aligned}$$

Theorem

Let $\lim_{k\to\infty} \{t^{(k)}\} = 0$*, let* $\boldsymbol{x}^{(k)}$ *be KKT points of NLP*$(t^{(k)})$ *with* $\lim_{k\to\infty} \{\boldsymbol{x}^{(k)}\} = \boldsymbol{x}^*$*, and let MPCC-MFCQ hold at* $\boldsymbol{x}^*$*. Then* $\boldsymbol{x}^*$ *is a C-stationary point.*

Lin-Fukushima Relaxation



Lin-Fukushima relaxation for the MPCC constraint $0 \leqslant u \perp v \geqslant 0$.

Steffensen-Ulbrich Smoothing-Relaxation

Solve a sequence of parameterized NLPs for $t \geqslant 0$:

(NLP(t))

$$\begin{aligned} \min_{\boldsymbol{x} \in \mathbb{R}^n} \ & F(\boldsymbol{x}) \\ \text{s.t.} \ & C(\boldsymbol{x}) = \mathbf{0}, \ D(\boldsymbol{x}) \geqslant \mathbf{0} \\ & G_i(\boldsymbol{x}) \geqslant 0, \ H_i(\boldsymbol{x}) \geqslant 0, \ 1 \leqslant i \leqslant c \\ & \Phi^{SU}(G_i(\boldsymbol{x}), H_i(\boldsymbol{x})) \leqslant 0, \ 1 \leqslant i \leqslant c \end{aligned}$$

wherein

$$\Phi^{SU}(u, v) = u + v - \phi_t(u - v)$$

and

$$\phi_t(a) = \begin{cases} |a| & \text{if } |a| \geqslant t \\ t\theta(a/t) & \text{if } |a| < t \end{cases}$$

and $\theta : (-1, 1) \to \mathbb{R}$ a certain regularization function.

Theorem

Let $\lim_{k\to\infty} \{t^{(k)}\} = 0$, let $\boldsymbol{x}^{(k)}$ be KKT points of NLP$(t^{(k)})$ with $\lim_{k\to\infty} \{\boldsymbol{x}^{(k)}\} = \boldsymbol{x}^*$, and let MPCC-CPLD hold at $\boldsymbol{x}^*$. Then $\boldsymbol{x}^*$ is a C-stationary point.

Kadrani's Kinked Relaxation

Solve a sequence of parameterized NLPs for $t \geqslant 0$:

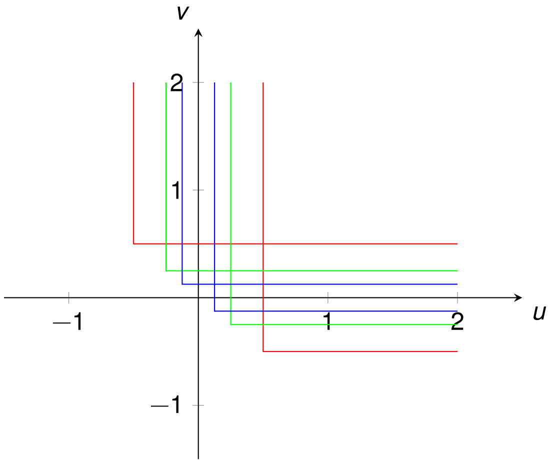
$$
(\text{NLP}(t)) \qquad \boxed{\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^n}\ & F(\boldsymbol{x}) \\ \text{s.t.}\ & C(\boldsymbol{x}) = \mathbf{0},\ D(\boldsymbol{x}) \geqslant \mathbf{0} \\ & G_i(\boldsymbol{x}) \geqslant -t,\ H_i(\boldsymbol{x}) \geqslant -t,\ 1 \leqslant i \leqslant c \\ & (G_i(\boldsymbol{x}) - t)(H_i(\boldsymbol{x}) - t) \leqslant 0,\ 1 \leqslant i \leqslant c \end{aligned}}
$$

This is not really a relaxation as it excludes the regions $[0, t) \times \{0\}$ and $\{0\} \times [0, t]$ and creates a disjoint feasible set. Nonetheless, it was the first relaxation for which one can show:

Theorem

Let $\lim_{k\to\infty} \{t^{(k)}\} = 0$, let $\boldsymbol{x}^{(k)}$ be KKT points of NLP$(t^{(k)})$ with $\lim_{k\to\infty} \{\boldsymbol{x}^{(k)}\} = \boldsymbol{x}^*$, and let MPCC-CPLD hold at $\boldsymbol{x}^*$. Then $\boldsymbol{x}^*$ is an M-stationary point.

Kadrani's Kinked Relaxation



Kandrani's relaxation for the MPCC constraint $0 \leqslant u \perp v \geqslant 0$.

Kanzow-Schwartz' Kinked Relaxation

Consider the NCP function

$$\phi(a,b) := \begin{cases} ab & \text{if } a+b \geqslant 0 \\ -\frac{1}{2}(a^2+b^2) & \text{if } a+b < 0 \end{cases}$$

Solve a sequence of parameterized NLPs for $t \geqslant 0$:

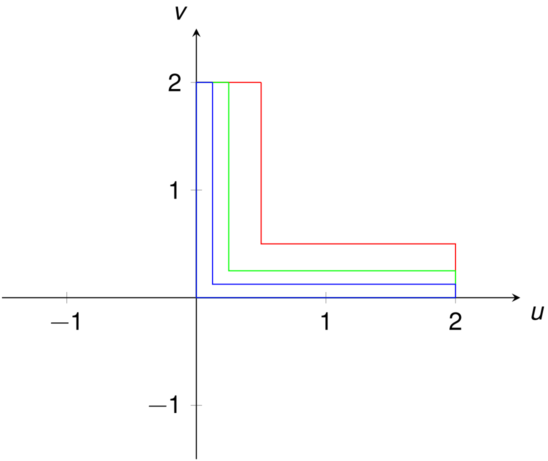
$$(\text{NLP}(t)) \qquad \boxed{\begin{array}{ll} \min\limits_{\boldsymbol{x}\in\mathbb{R}^n} & F(\boldsymbol{x}) \\ \text{s.t.} & C(\boldsymbol{x}) = \mathbf{0},\ D(\boldsymbol{x}) \geqslant \mathbf{0} \\ & G_i(\boldsymbol{x}) \geqslant 0,\ H_i(\boldsymbol{x}) \geqslant 0,\ 1 \leqslant i \leqslant c \\ & \phi(G_i(\boldsymbol{x}) - t, H_i(\boldsymbol{x}) - t) \leqslant 0,\ 1 \leqslant i \leqslant c \end{array}}$$

This has been derived from Kadrani's formulation and addresses the disjointness issue.

Theorem

Let $\lim\limits_{k\to\infty}\{t^{(k)}\} = 0$, let $\boldsymbol{x}^{(k)}$ be KKT points of NLP$(t^{(k)})$ with $\lim\limits_{k\to\infty}\{\boldsymbol{x}^{(k)}\} = \boldsymbol{x}^*$, and let MPCC-CPLD hold at $\boldsymbol{x}^*$. Then $\boldsymbol{x}^*$ is an M-stationary point.

Kanzow-Schwartz Kinked Relaxation



The Kanzow-Schwartz relaxation for the MPCC constraint $0 \leqslant u \perp v \geqslant 0$.

Inexactness Effects

Relaxation and smoothing convergence theorems assume that subproblems are solved exactly. Computing inexact KKT points with tolerance $0 < \varepsilon^k$ has detrimental effects on their validity.

- Scholtes, Lin-Fukushima: If $\varepsilon^k \in O(t_k)$ then x^* is C-stationary.
- All others: x^* will only be weakly stationary!

Example (Kanzow-Schwartz):

$$\min_{\boldsymbol{x} \in \mathbb{R}^2} -x_1 - x_2 \text{ s.t. } 0 \leqslant x_1 \perp x_2 \geqslant 0.$$

For a sequence $t^k \to 0$ and assuming $\varepsilon^k = (t^k)^2$, one verifies the family of ε^t-KKT points

$$x^t = ((1-t)t, (1-t)t)^T$$

for NLP(t^k) with multiplier $\delta^t = 1/\varepsilon_t$ for the NCP inequality of the Kanzow-Schwartz relaxation.

The limit $(0, 0)$ is only C-stationary (and satisfies MPCC-LICQ). This is in contrast to the theorem guaranteeing M-stationarity.

Penalty approach

A generic and common approach to deal with difficult constraints C in an NLP

$$\min_{\boldsymbol{x} \in \mathbb{R}^n} F(\boldsymbol{x}) \text{ s.t. } C(\boldsymbol{x}) = \mathbf{0},\ x \in \Omega$$

is to replace them by a suitable penalty term $P_C(\boldsymbol{x})$ in the objective,

$$\min_{\boldsymbol{x} \in \mathbb{R}^n} F(\boldsymbol{x}) + \rho P_C(\boldsymbol{x}) \text{ s.t. } x \in \Omega,$$

which one tries to drive to zero by taking $\rho \to \infty$.

Exact penalty functions: $P(\boldsymbol{x}) = 0$ iff $\boldsymbol{x}$ solves the original problem.

ℓ_1 and ℓ_∞ penalty functions are exact, but non-smooth.

The ℓ_2 penalty function is not exact, but differentiable.

ℓ_1-penalty approach

The ℓ_1 penalty approach for MPCCs in vertical form reads as follows.
For $\rho \gg 0$ solve

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^n} \; & F(\boldsymbol{x}) + \rho \sum_{i=1}^{c} |u_i v_i| \\ \text{s.t. } & \mathbf{0} \leqslant \boldsymbol{u} = G(\boldsymbol{x}) \\ & \mathbf{0} \leqslant \boldsymbol{v} = H(\boldsymbol{x}) \end{aligned}$$

This is non-smooth, but may also be written as

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^n} \; & F(\boldsymbol{x}) + \rho \sum_{i=1}^{c} \xi_i \\ \text{s.t. } & u_i v_i \leqslant \xi_i, \; 1 \leqslant i \leqslant c \\ & \mathbf{0} \leqslant \boldsymbol{u} = G(\boldsymbol{x}) \\ & \mathbf{0} \leqslant \boldsymbol{v} = H(\boldsymbol{x}) \end{aligned}$$

Assuming G and H are sufficiently regular, this violates CQs only in $\xi = 0$.

ℓ_∞-penalty approach

The ℓ_∞ penalty approach for MPCCs in vertical form reads as follows:
For $\rho \gg 0$ solve

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^n} \; & F(\boldsymbol{x}) + \rho \max_{1\leqslant i\leqslant c}\{|u_i v_i|\} \\ \text{s.t. } & \boldsymbol{0} \leqslant \boldsymbol{u} = G(\boldsymbol{x}) \\ & \boldsymbol{0} \leqslant \boldsymbol{v} = H(\boldsymbol{x}) \end{aligned}$$

This is non-smooth, but may also be written as

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^n} \; & F(\boldsymbol{x}) + \rho\xi \\ \text{s.t. } & u_i v_i \leqslant \xi, \; 1 \leqslant i \leqslant c \\ & \boldsymbol{0} \leqslant \boldsymbol{u} = G(\boldsymbol{x}) \\ & \boldsymbol{0} \leqslant \boldsymbol{v} = H(\boldsymbol{x}) \end{aligned}$$

Assuming G and H are regular themselves, this violates CQs in $\xi = 0$.

Interior Point Methods (refresher)

The nonlinear program

$$\min_{x \in \mathbb{R}^n} F(x) \text{ s.t } H(x) \geqslant 0$$

is reformulated using a log-barrier,

$$\min_{\boldsymbol{x} \in \mathbb{R}^n,\ \boldsymbol{w} \in \mathbb{R}^k} F(\boldsymbol{x}) - \theta \sum_{i=1}^{k} \log(w_i)$$
$$\text{s.t. } H(\boldsymbol{x}) - \boldsymbol{w} = \mathbf{0}$$

The first order necessary optimality conditions read

$$\begin{aligned} \nabla F(\boldsymbol{x}) - \nabla H(\boldsymbol{x})\boldsymbol{\lambda} &= \mathbf{0} \\ -\theta\mathbf{1} + W\Lambda\mathbf{1} &= \mathbf{0} \\ H(\boldsymbol{x}) - \boldsymbol{w} &= \mathbf{0} \end{aligned}$$

Interior Point Methods (refresher)

Newton's method applied to this root finding problem solves

$$\begin{pmatrix} -\nabla^2_{\boldsymbol{xx}}\mathcal{L}(\boldsymbol{x},\boldsymbol{\lambda}) & \nabla H(\boldsymbol{x}) \\ \nabla H(\boldsymbol{x})^T & W\Lambda^{-1} \end{pmatrix} \begin{pmatrix} \Delta\boldsymbol{x} \\ \Delta\boldsymbol{\lambda} \end{pmatrix} \begin{pmatrix} \sigma \\ \rho + W\Lambda^{-1}\gamma \end{pmatrix},$$

wherein

$$\begin{aligned} \sigma &:= \nabla F(\boldsymbol{x}) - \nabla H(\boldsymbol{x})\boldsymbol{\lambda}, \\ \gamma &:= \theta W^{-1}\mathbf{1} - \Lambda, \\ \rho &:= \boldsymbol{w} - H(\boldsymbol{x}). \end{aligned}$$

and

$$\Delta\boldsymbol{w} = W\Lambda^{-1}(\gamma - \Delta\boldsymbol{\lambda}).$$

Step sizes for $\boldsymbol{w}$ and $\boldsymbol{\lambda}$ are chosen such that they remain positive and reduce either the barrier or the infeasibility without increasing the respective other quantity by too much.

Interior Point Methods for MPCC

Remember GCQ implies MPCC-LICQ. The first admits unbounded KKT multipliers while the second is already considered restrictive. Hence,

$$W\Lambda = \theta \mathbf{1}$$

in the IP method will lead to slack entries $w_i \to 0$ as $\lambda_i \to \infty$. The IP method has to pick tiny step sizes as a consequence. If $\theta \gg 0$ when this happens, the IP method may stall.

Consider the log-barrier formulation of the MPCC:

$$\begin{aligned}
\min_{\boldsymbol{x}\in\mathbb{R}^n} \; & F(\boldsymbol{x}) - \theta \sum_{i=1}^{c} \log w_i - \theta \sum_{i=1}^{c} \log t_i - \theta \sum_{i=1}^{c} \log z_i \\
\text{s.t. } & G(\boldsymbol{x}) - \boldsymbol{u} = \mathbf{0} \\
& H(\boldsymbol{x}) - \boldsymbol{v} = \mathbf{0} \\
& \boldsymbol{u} \circ \boldsymbol{v} + \boldsymbol{w} = \mathbf{0} \\
& \boldsymbol{u} - \boldsymbol{t} = \mathbf{0} \\
& \boldsymbol{v} - \boldsymbol{z} = \mathbf{0}
\end{aligned}$$

Interior Point Methods for MPCC

We partition $u = (u_I, u_J)$ and $v = (v_I, v_J)$ such that $u_I = 0$ and $v_J = 0$. The stationarity part of the KKT conditions for this log-barrier problem reads

$$\nabla_{\boldsymbol{u}}\mathcal{L} = -\lambda^G + \begin{pmatrix} 0 & 0 \\ 0 & V_J \end{pmatrix} \begin{pmatrix} \lambda_I^w \\ \lambda_J^w \end{pmatrix} + \begin{pmatrix} \lambda_I^t \\ \lambda_J^t \end{pmatrix}$$

$$\nabla_{\boldsymbol{v}}\mathcal{L} = -\lambda^H + \begin{pmatrix} U_I & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \lambda_I^w \\ \lambda_J^w \end{pmatrix} + \begin{pmatrix} \lambda_I^z \\ \lambda_J^z \end{pmatrix}$$

$$\nabla_{w_i}\mathcal{L} = 0 \text{ implies } \lambda_i^w w_i = \theta$$

$$\nabla_{z_i}\mathcal{L} = 0 \text{ implies } \lambda_i^z z_i = \theta$$

$$\nabla_{t_i}\mathcal{L} = 0 \text{ implies } \lambda_i^t t_i = \theta$$

Even under LLSCC, $\theta \to 0$ will result in $\lambda_i^w \to \infty$ or $\lambda_i^z \to \infty$ for some components.

If a complementarity pair is non-strict, $u_i = v_i = 0$, then the corresponding λ_i^w drops out and the system becomes rank-deficient.

Theorem

If an IP method is applied to the ℓ_1- or ℓ_∞-penalty form of an MPCC that satisfies MPCC-LICQ, the set of classical KKT multipliers remains bounded.

Sequential Quadratic Programming (refresher)

The nonlinear program

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} F(\boldsymbol{x}) \text{ s.t } H(\boldsymbol{x}) \geqslant \mathbf{0}$$

is solved by iteratively minimizing a quadratic model of the objective and a linear model of the constraints:

(QP^k)

$$\begin{aligned} \min_{\Delta\boldsymbol{x}\in\mathbb{R}^n} \ & \tfrac{1}{2}\Delta\boldsymbol{x}^T B^k \Delta\boldsymbol{x} + \nabla F(\boldsymbol{x}^k)^T \Delta\boldsymbol{x} \\ \text{s.t } & \nabla H(\boldsymbol{x}^k)^T \Delta\boldsymbol{x} + H(\boldsymbol{x}^k) \geqslant \mathbf{0} \end{aligned}$$

wherein $B^k \approx \nabla^2_{xx}\mathcal{L}(\boldsymbol{x}^k, \boldsymbol{\lambda}^k)$.
Step sizes for $\Delta\boldsymbol{x}$ are chosen such that, for example, they reduce a merit function balancing the objective against the infeasibility.

Sequential Quadratic Programming

Consider now the application of SQP to the MPCC.

The quadratic model reads

$$(\mathrm{QP}^k) \qquad \min_{\Delta \boldsymbol{x} \in \mathbb{R}^n} \tfrac{1}{2} \Delta \boldsymbol{x}^T B^k \Delta \boldsymbol{x} + \nabla F(\boldsymbol{x}^k)^T \Delta \boldsymbol{x}$$

$$\text{s.t } \left(H_i(\boldsymbol{x}^k) \nabla G_i(\boldsymbol{x}^k)^T + G_i(\boldsymbol{x}^k) \nabla H_i(\boldsymbol{x}^k)^T \right) \Delta \boldsymbol{x} = 0, \ 1 \leqslant i \leqslant c$$

In a strictly complementary pair, $G_i(\boldsymbol{x}^k) = H_i(\boldsymbol{x}^k) = 0$, the linearization drops out and the step $\Delta \boldsymbol{x}$ can point into arbitrary infeasible direction w.r.t. the complementarity indexed by i. This leads to near-zero step size and stalling of the method.

Linearizations of the MPCC constraints $\mathbf{0} \leqslant \boldsymbol{u} \perp \boldsymbol{v} \geqslant \mathbf{0}$ are not sensible.

Sequential Quadratic Programming

SQP applied to the ℓ_1-penalty formulation of MPCC.

Refresher: The ℓ_1-penalty formulation of an MPCC in vertical form reads

$$\begin{aligned}\min_{\substack{\boldsymbol{x}\in\mathbb{R}^n,\boldsymbol{u}\in\mathbb{R}^c\\ \boldsymbol{v}\in\mathbb{R}^c,\boldsymbol{\xi}\in\mathbb{R}^c}} \quad & F(\boldsymbol{x}) + \rho\sum_{i=1}^{c}\xi_i \\ \text{s.t. } & u_i v_i \leqslant \xi_i, \ 1 \leqslant i \leqslant c \\ & \mathbf{0} \leqslant \boldsymbol{u} = G(\boldsymbol{x}) \\ & \mathbf{0} \leqslant \boldsymbol{v} = H(\boldsymbol{x})\end{aligned}$$

Sequential Quadratic Programming

The quadratic model reads

$$\begin{aligned}
\min_{\substack{\Delta \boldsymbol{x}\in\mathbb{R}^n, \Delta \boldsymbol{u}\in\mathbb{R}^c \\ \Delta \boldsymbol{v}\in\mathbb{R}^c, \Delta \boldsymbol{\xi}\in\mathbb{R}^c}} \quad & \tfrac{1}{2}\Delta \boldsymbol{x}^T B^k \Delta \boldsymbol{x} + \nabla F(\boldsymbol{x}^k)^T \Delta \boldsymbol{x} + \rho \sum_{i=1}^{c} \Delta \xi_i \\
\text{s.t} \quad & v_i^k \Delta u_i + u_i^k \Delta v_i - \Delta \xi_i - \xi_i^k \leqslant 0,\ 1 \leqslant i \leqslant c \\
& \nabla G_i(\boldsymbol{x}^k)^T \Delta \boldsymbol{x} + G_i(\boldsymbol{x}^k) - \Delta u_i = 0,\ 1 \leqslant i \leqslant c \\
& \nabla H_i(\boldsymbol{x}^k)^T \Delta \boldsymbol{x} + H_i(\boldsymbol{x}^k) - \Delta v_i = 0,\ 1 \leqslant i \leqslant c \\
& \Delta \boldsymbol{u} + \boldsymbol{u}^k \geqslant \boldsymbol{0} \\
& \Delta \boldsymbol{v} + \boldsymbol{v}^k \geqslant \boldsymbol{0}
\end{aligned}$$

In a strictly complementary pair $u_i^k v_i^k = 0$, the linearization will result in $\Delta \xi_i^* = -\xi_i^k$. The unknowns $(\Delta u_i, \Delta v_i)$ drop out, but full row rank of the reduced linear system may in general be maintained by an active-set method for solving the QP.

This usually results in $(\Delta u_i^*, \Delta v_i^*) = (0, 0)$ and attracts non-strict local solutions.

Augmented Lagrangian Methods (refresher)

The nonlinear program

$$\min_{\boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{w} \in \mathbb{R}^k} F(\boldsymbol{x}) \text{ s.t } H(\boldsymbol{x}) - \boldsymbol{w} = \boldsymbol{0}$$

is solved by iteratively solving the bound-constrained problems

$$(\text{NLP}(\lambda^k, \mu^k)) \qquad \min_{\boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{w} \in \mathbb{R}^k} F(\boldsymbol{x}) + \frac{\mu^k}{2} \sum_{i=1}^{k} w_i^2 + \sum_{i=1}^{k} \lambda_i^k w_i \quad \text{s.t. } \boldsymbol{w} \geqslant \boldsymbol{0}$$

and updating

$$\lambda^{k+1} \leftarrow \lambda^k + \mu^k \boldsymbol{w}.$$

The penalty μ^k is adjusted by monitoring the infeasibility. The subproblems NLP(λ^k, μ^k) can be solved using, e.g., SQP with BQP subproblems.

Augmented Lagrangian Methods

Application to the MPCC in vertical form:

$$\min_{\substack{\boldsymbol{x}\in\mathbb{R}^n,\boldsymbol{u}\in\mathbb{R}^c\\ \boldsymbol{v}\in\mathbb{R}^c}} F(\boldsymbol{x}) \text{ s.t. } \boldsymbol{u}\circ\boldsymbol{v}=\boldsymbol{0}$$
$$\boldsymbol{u}=G(\boldsymbol{x}),\ \ \boldsymbol{v}=H(\boldsymbol{x}),$$
$$\boldsymbol{u}\geqslant\boldsymbol{0},\ \boldsymbol{v}\geqslant\boldsymbol{0}.$$

Iteratively solve the bound-constrained problems NLP(λ^k, μ^k)

$$\min_{\substack{\boldsymbol{x}\in\mathbb{R}^n,\boldsymbol{u}\in\mathbb{R}^c\\ \boldsymbol{v}\in\mathbb{R}^c}} F(\boldsymbol{x}) + \frac{\mu^k}{2}\sum_{i=1}^{c}\left[(G_i(\boldsymbol{x})-u_i)^2+(H_i(\boldsymbol{x})-v_i)^2+(u_iv_i)^2\right]$$
$$+\sum_{i=1}^{c}\lambda_i^k\left[(G_i(\boldsymbol{x})-u_i)+(H_i(\boldsymbol{x})-v_i)+(u_iv_i)\right]$$
$$\text{s.t. } \boldsymbol{u}\geqslant\boldsymbol{0},\ \ \boldsymbol{v}\geqslant\boldsymbol{0}.$$

and updating

$$\lambda^{k+1}\leftarrow\lambda^k+\mu^k\begin{pmatrix}G(\boldsymbol{x})-\boldsymbol{u}\\ H(\boldsymbol{x})-\boldsymbol{v}\\ \boldsymbol{u}\circ\boldsymbol{v}\end{pmatrix}.$$

Non-smooth subproblems

So far we relaxed / smoothed / regularized the non-smooth MPCC in the original problem and generated smooth / linear subproblems.

Another approach is to keep non-differentiable structures in the subproblems generated by iterative methods:

- Linear systems with complementarity constraints: LCPs
- LPs with complementarity constraints: LPCCs
- Bound constrained QPs with complementarity constraints: BQPCCs
- QPs with complementarity constraints: QPCCs

With the exception of LCPs, it is somehwhat difficult to find solvers for these, let alone reliable and fast ones.

Sequential QPCC

Sequential quadratic programming with linearization applied to the two branches of every complementarity constraint.

The quadratic model with linearized complementarity constraints (QPCC) reads

$$(\text{QPCC}^k) \qquad \min_{\Delta\boldsymbol{x}\in\mathbb{R}^n} \tfrac{1}{2}\Delta\boldsymbol{x}^T B^k \Delta\boldsymbol{x} + \nabla F(\boldsymbol{x}^k)^T \Delta\boldsymbol{x}$$

$$\text{s.t } 0 \leqslant \nabla G(\boldsymbol{x}^k)\Delta\boldsymbol{x} + G(\boldsymbol{x}^k) \perp \nabla H(\boldsymbol{x}^k)\Delta\boldsymbol{x} + H(\boldsymbol{x}^k) \geqslant 0$$

This requires an algorithm for solving QPCCs.

Convergence to C-stationary points is all one can show.

Sequential QPCC

Failure of SQPCC: Consider the problem

$$\min_{\boldsymbol{x}\in\mathbb{R}^2} (x_1-1)^2 + x_2^3 + x_2^2 \text{ s.t. } 0 \leqslant x_1 \perp x_2 \geqslant 0.$$

The point $\hat{\boldsymbol{x}}(0,0)^T$ is M-stationary with trivial descent direction $(1,0)^T$ to $\boldsymbol{x}^* = (1,0)$, which is B-stationary but not S-stationary. Compute

$$\nabla f(\boldsymbol{x}) = \begin{pmatrix} 2(x_1-1) \\ 2x_2^2+2x_2 \end{pmatrix}, \; \nabla^2\mathcal{L}(\boldsymbol{x}) = \begin{pmatrix} 2 & 0 \\ 0 & 6x_2+2 \end{pmatrix}.$$

The QPEC at any $x^{(k)}$ is ($x_1^{(k)} = 0$)

$$\min_{\boldsymbol{d}} \; d_1^2 + (3x_2^{(k)}+2)d_2^2 + 2(x_1^{(k)}-1)d_1 + (3{x_2^{(k)}}^2 + 2x_2^{(k)})d_2 \text{ s.t. } 0 \leqslant d_1 \perp x_2^{(k)} + d_2 \geqslant 0.$$

Starting in $x^{(0)} = (0,t)$ for some $t \in (0,1)$ SQPEC will generate the sequence

$$x^{(k)} = \left(0, \; \frac{3{y^{(k)}}^2}{6y^{(k)}+2}\right)^T$$

which shows q-quadratic convergence (as desired) but to $(0,0)$, which is M-stationary.

Sequential LPCC Programming (SLPCC)

Instead of solving QPs to compute Newton-type steps, we can solve LPs and compute constrained gradient-descent steps.

Sequential linear programming with linearization applied to the two branches of every complementarity constraint.

The linear model with linearized complementarity constraints (LPCC) reads

$$
\begin{aligned}
(\text{LPCC}^k(\Delta^k)) \quad & \min_{\Delta \boldsymbol{x} \in \mathbb{R}^n} \nabla F(\boldsymbol{x}^k)^T \Delta \boldsymbol{x} \\
& \text{s.t } 0 \leqslant \nabla G(\boldsymbol{x}^k)\Delta \boldsymbol{x} + G(\boldsymbol{x}^k) \perp \nabla H(\boldsymbol{x}^k)\Delta \boldsymbol{x} + H(\boldsymbol{x}^k) \geqslant 0 \\
& \quad \|\Delta \boldsymbol{x}\|_\infty \leqslant \Delta_{\text{LP}}^k
\end{aligned}
$$

This requires an additional trust-region constraint for boundedness from below, and handling the trust region in the neighborhood of the complementarity kink requires special attention.

Also, this requires an algorithm for solving LPCCs. Also, do this for the vertical form.

Convergence to B-stationary points can be shown assuming LPCC can be solved.

SLPCC with an Accelerating Newton Step

Similar to SLP-EQP methods in NLP. Solve

$$(\text{LPCC}(\Delta^k_{\text{LP}})) \quad \min_{\Delta \boldsymbol{x} \in \mathbb{R}^n} \nabla F(\boldsymbol{x}^k)^T \Delta \boldsymbol{x}$$

$$\text{s.t } 0 \leqslant \nabla G(\boldsymbol{x}^k)\Delta \boldsymbol{x} + G(\boldsymbol{x}^k) \perp \nabla H(\boldsymbol{x}^k)\Delta \boldsymbol{x} + H(\boldsymbol{x}^k) \geqslant 0$$

$$\|\Delta \boldsymbol{x}\|_\infty \leqslant \Delta^k_{\text{LP}}$$

to find a constrained gradient-descent step Δx_{LP} onto a new active set $\mathcal{A}$. Fix this and solve

$$(\text{EQP}^k) \quad \min_{\Delta \boldsymbol{x} \in \mathbb{R}^n} \tfrac{1}{2} \Delta \boldsymbol{x}^T B^k \Delta \boldsymbol{x} + \nabla F(\boldsymbol{x}^k)^T \Delta \boldsymbol{x}$$

$$\text{s.t } 0 = \nabla G_i(\boldsymbol{x}^k)\Delta \boldsymbol{x} + G_i(\boldsymbol{x}^k), \ i \in \mathcal{A} \cap I_{0+}$$

$$0 = \nabla H(\boldsymbol{x}^k)\Delta \boldsymbol{x} + H(\boldsymbol{x}^k), \ i \in \mathcal{A} \cap I_{+0}$$

for a Newton-type step step Δx_{EQP} on the null-space of the constraints in $\mathcal{A}$. This is just an indefinite linear system if B^k positive semidefinite. Requires an ℓ_2 trust-region if B^k indefinite $\rightarrow$ CG-Steihaug.

Take a convex combination of these steps that reduces a merit function.

Augmented Lagrangian Methods

Application to the MPCC in vertical form:

$$\min_{\substack{\boldsymbol{x}\in\mathbb{R}^n,\boldsymbol{u}\in\mathbb{R}^c\\ \boldsymbol{v}\in\mathbb{R}^c}} F(\boldsymbol{x}) \text{ s.t. } 0 \leqslant \boldsymbol{u} \perp \boldsymbol{v} \geqslant \boldsymbol{0}$$
$$\boldsymbol{u} = G(\boldsymbol{x}),\ \ \boldsymbol{v} = H(\boldsymbol{x}),$$
$$\boldsymbol{u} \geqslant \boldsymbol{0}, \boldsymbol{v} \geqslant \boldsymbol{0}.$$

Iteratively solve the bound-constrained complementarity problems MPCC(λ^k, μ^k)

$$\min_{\substack{\boldsymbol{x}\in\mathbb{R}^n,\boldsymbol{u}\in\mathbb{R}^c\\ \boldsymbol{v}\in\mathbb{R}^c}} F(\boldsymbol{x}) + \frac{\mu^k}{2}\sum_{i=1}^{c}\left[(G_i(\boldsymbol{x}) - u_i)^2 + (H_i(\boldsymbol{x}) - v_i)^2\right]$$
$$+ \sum_{i=1}^{c} \lambda_i^k \left[(G_i(\boldsymbol{x}) - u_i) + (H_i(\boldsymbol{x}) - v_i)\right]$$
$$\text{s.t. } \boldsymbol{0} \leqslant \boldsymbol{u} \perp \boldsymbol{v} \geqslant 0.$$

and update

$$\lambda^{k+1} \leftarrow \lambda^k + \mu^k \begin{pmatrix} G(\boldsymbol{x}) - \boldsymbol{u} \\ H(\boldsymbol{x}) - \boldsymbol{v} \end{pmatrix}.$$

Solving MPCC (λ^k, μ^k) can be done by trust region SQP. The resulting BQPCCs enjoy MPCC-LICQ and can be solved efficiently by projected gradient descent with CG acceleration.