5. MPECs: nonsmooth modelling and optimality conditions

Armin Nurkanović

Systems Control and Optimization Laboratory (syscop), University of Freiburg, Germany

Winter School on Numerical Methods for Optimal Control of Nonsmooth Systems
École des Mines de Paris
February 3-5, 2025, Paris, France

universität freiburg

Outline of the lecture

Mathematical programs with complemenatarities constraints (MPCC)

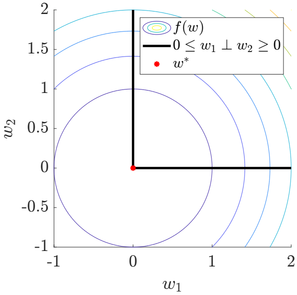
MPCC

$$\min_{w \in \mathbb{R}^n} \quad f(w) \tag{1a}$$

$$\text{s.t.} \quad g(w) = 0, \tag{1b}$$

$$h(w) \geq 0, \tag{1c}$$

$$0 \leq G(w) \perp H(w) \geq 0. \tag{1d}$$

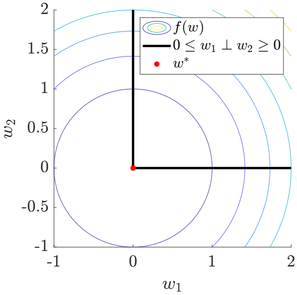


$$\min_{w \in \mathbb{R}^n} w_1^2 + w_2^2 \quad \text{s.t.} \quad 0 \leq w_1 \perp w_2 \geq 0$$

Mathematical programs with complemenatarities constraints (MPCC)

MPCC

$$\min_{w \in \mathbb{R}^n} \quad f(w) \tag{1a}$$
$$\text{s.t.} \quad g(w) = 0, \tag{1b}$$
$$h(w) \geq 0, \tag{1c}$$
$$0 \leq G(w) \perp H(w) \geq 0. \tag{1d}$$



$$\min_{w \in \mathbb{R}^n} w_1^2 + w_2^2 \quad \text{s.t.} \quad 0 \leq w_1 \perp w_2 \geq 0$$

Reformulated into an nonlinear program (NLP) by replacing (1d) by:

1. $G(w) \geq 0,\ H(w) \geq 0,\ G_i(w)H_i(w) \leq 0,\ i = 1, \dots, m,$

Mathematical programs with complemenatarities constraints (MPCC)

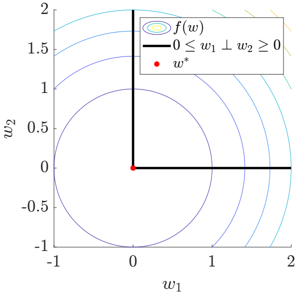
MPCC

$$\min_{w\in\mathbb{R}^n} \quad f(w) \tag{1a}$$

$$\text{s.t.} \quad g(w) = 0, \tag{1b}$$

$$h(w) \geq 0, \tag{1c}$$

$$0 \leq G(w) \perp H(w) \geq 0. \tag{1d}$$



$$\min_{w\in\mathbb{R}^n} w_1^2 + w_2^2 \quad \text{s.t.} \quad 0 \leq w_1 \perp w_2 \geq 0$$

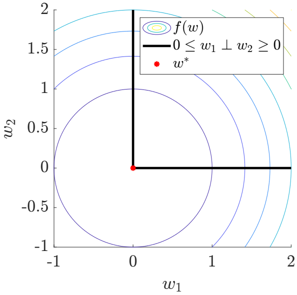
Reformulated into an nonlinear program (NLP) by replacing (1d) by:

1. $G(w) \geq 0,\ H(w) \geq 0,\ G_i(w)H_i(w) \leq 0,\ i = 1, \dots, m,$
2. $G(w) \geq 0,\ H(w) \geq 0,\ G(w)^\top H(w) \leq 0,$

Mathematical programs with complemenatarities constraints (MPCC)

MPCC

$$\min_{w \in \mathbb{R}^n} \quad f(w) \tag{1a}$$
$$\text{s.t.} \quad g(w) = 0, \tag{1b}$$
$$h(w) \geq 0, \tag{1c}$$
$$0 \leq G(w) \perp H(w) \geq 0. \tag{1d}$$



$$\min_{w \in \mathbb{R}^n} w_1^2 + w_2^2 \quad \text{s.t.} \quad 0 \leq w_1 \perp w_2 \geq 0$$

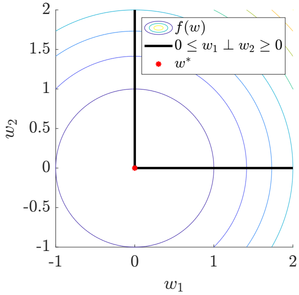
Reformulated into an nonlinear program (NLP) by replacing (1d) by:

1. $G(w) \geq 0, \; H(w) \geq 0, \; G_i(w)H_i(w) \leq 0, \; i = 1, \ldots, m,$
2. $G(w) \geq 0, \; H(w) \geq 0, \; G(w)^\top H(w) \leq 0,$
3. $G(w) \geq 0, \; H(w) \geq 0, \; G(w)^\top H(w) = 0,$

Mathematical programs with complemenatarities constraints (MPCC)

MPCC

$$\min_{w \in \mathbb{R}^n} \quad f(w) \tag{1a}$$
$$\text{s.t.} \quad g(w) = 0, \tag{1b}$$
$$h(w) \geq 0, \tag{1c}$$
$$0 \leq G(w) \perp H(w) \geq 0. \tag{1d}$$



Reformulated into an nonlinear program (NLP) by replacing (1d) by:

1. $G(w) \geq 0, \; H(w) \geq 0, \; G_i(w)H_i(w) \leq 0, \; i = 1, \ldots, m,$
2. $G(w) \geq 0, \; H(w) \geq 0, \; G(w)^\top H(w) \leq 0,$
3. $G(w) \geq 0, \; H(w) \geq 0, \; G(w)^\top H(w) = 0,$
4. $G(w) \geq 0, \; H(w) \geq 0, \; G_i(w)H_i(w) = 0, \; i = 1, \ldots, m,$

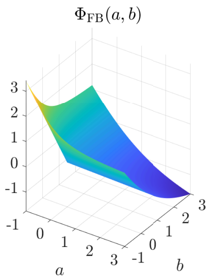
MPCC

$$\min_{w \in \mathbb{R}^n} \quad f(w) \tag{1a}$$

$$\text{s.t.} \quad g(w) = 0, \tag{1b}$$

$$h(w) \geq 0, \tag{1c}$$

$$0 \leq G(w) \perp H(w) \geq 0. \tag{1d}$$



$$\Phi_{\mathrm{FB}}(a,b) = \sqrt{a^2 + b^2} - a - b$$

Reformulated into an nonlinear program (NLP) by replacing (1d) by:

1. $G(w) \geq 0,\ H(w) \geq 0,\ G_i(w)H_i(w) \leq 0,\ i = 1, \ldots, m,$
2. $G(w) \geq 0,\ H(w) \geq 0,\ G(w)^\top H(w) \leq 0,$
3. $G(w) \geq 0,\ H(w) \geq 0,\ G(w)^\top H(w) = 0,$
4. $G(w) \geq 0,\ H(w) \geq 0,\ G_i(w)H_i(w) = 0,\ i = 1, \ldots, m,$
5. $\Phi_{\mathrm{C}}(G(w), H(w)) = 0.$

Mathematical programs with complemenatarities constraints (MPCC)

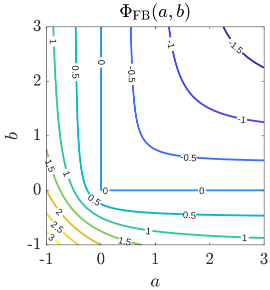
MPCC

$$\min_{w \in \mathbb{R}^n} \quad f(w) \tag{1a}$$

$$\text{s.t.} \quad g(w) = 0, \tag{1b}$$

$$h(w) \geq 0, \tag{1c}$$

$$0 \leq G(w) \perp H(w) \geq 0. \tag{1d}$$



$$\Phi_{\mathrm{FB}}(a,b) = \sqrt{a^2 + b^2} - a - b$$

Reformulated into an nonlinear program (NLP) by replacing (1d) by:

1. $G(w) \geq 0,\ H(w) \geq 0,\ G_i(w)H_i(w) \leq 0,\ i = 1, \ldots, m,$
2. $G(w) \geq 0,\ H(w) \geq 0,\ G(w)^\top H(w) \leq 0,$
3. $G(w) \geq 0,\ H(w) \geq 0,\ G(w)^\top H(w) = 0,$
4. $G(w) \geq 0,\ H(w) \geq 0,\ G_i(w)H_i(w) = 0,\ i = 1, \ldots, m,$
5. $\Phi_{\mathrm{C}}(G(w), H(w)) = 0.$

Vertical form of an MPCC

Usually, we keep the complementarity constraints as simple as possible:

Generic MPCC

$$\begin{aligned}
\min_{w \in \mathbb{R}^n} \quad & f(w) \\
\text{s.t.} \quad & g(w) = 0, \\
& h(w) \geq 0, \\
& 0 \leq G(w) \perp H(w) \geq 0.
\end{aligned}$$

Vertical form of an MPCC

Usually, we keep the complementarity constraints as simple as possible:

Vertical form MPCC

$$
\begin{aligned}
\min_{\substack{w \in \mathbb{R}^n \\ w_1, w_2 \in \mathbb{R}^m}} \quad & f(w) \\
\text{s.t.} \quad & g(w) = 0, \\
& h(w) \geq 0, \\
& w_1 - G(w) = 0, \\
& w_2 - H(w) = 0, \\
& 0 \leq w_1 \perp w_2 \geq 0.
\end{aligned}
$$

Vertical form of an MPCC

Usually, we keep the complementarity constraints as simple as possible:

Vertical form MPCC

$$
\begin{aligned}
\min_{w \in \mathbb{R}^n} \quad & f(w) \\
\text{s.t.} \quad & g(w) = 0, \\
& h(w) \geq 0, \\
& 0 \leq w_1 \perp w_2 \geq 0.
\end{aligned}
$$

where $w = (w_0, w_1, w_2) \in \mathbb{R}^n,\ w_0 \in \mathbb{R}^p,\ w_1, w_2 \in \mathbb{R}^m$, $n = p + 2m$

MPCC vs MPEC

Mathematical programs with equilibrium constraints (MPECs)

$$
\begin{aligned}
\min_{w \in \mathbb{R}^n} \quad & f(w) \\
\text{s.t.} \quad & w \in \mathcal{U} \\
& w_1 = \arg\min_{\hat{w}_1} \; F(\hat{w}_1; w_0) \text{ s.t. } \hat{w}_1 \in \mathcal{L}(w_0)
\end{aligned}
$$

- Instead of the lower level optimization problem, we may write a variational inequality, i.e., an *equilibirum constraints*.

MPCC vs MPEC

MPCC as MPEC

$$\begin{aligned} \min_{w \in \mathbb{R}^n} \quad & f(w) \\ \text{s.t.} \quad & g(w) = 0, \\ & h(w) \geq 0, \\ & w_1 \geq 0, \\ & w_2 = \arg\min_{\hat{w}_2} \; w_1^\top \hat{w}_2 \text{ s.t. } \hat{w}_2 \geq 0 \end{aligned}$$

- Instead of the lower level optimization problem, we may write a variational inequality, i.e., an *equilibirum constraints*.
- Because of the easier pronunciation, we often say MPEC but we mean MPCC.
- MPECs are more general than MPCCs, e.g., if the lower-level problem is nonconvex, then an MPCC is a relaxation of the MPEC.

An incomplete list of MPEC applications

- optimal control of hybrid/nonsmooth constraints (this winter school) [Baumrucker and Biegler, 2009, Guo and Ye, 2016, Vieira et al., 2019, Nurkanović, 2023]
- optimization with piecewise smooth functions, abs-normal froms [Hegerhorst-Schultchen et al., 2020]
- bi-level optimization (if the lower level problem is convex) [Kim et al., 2020]
- modeling of logical constraints [Pozharskiy et al., 2024]
- inverse optimization [Albrecht and Ulbrich, 2017, Hu et al., 2012]
- process and chemical engineering [Baumrucker et al., 2008, Biegler, 2010]
- robotics [Wensing et al., 2023]
- district heating networks [Krug et al., 2021]

Some literature on MPECs I would recommend: [Luo et al., 1996, Kim et al., 2020, Hu et al., 2012, Scheel and Scholtes, 2000, Flegel and Kanzow, 2005, Fletcher and Leyffer, 2004, Nurkanović et al., 2024, Kanzow and Schwartz, 2015, Hoheisel et al., 2013, Biegler, 2010, Hu et al., 2012]

Outline of the lecture

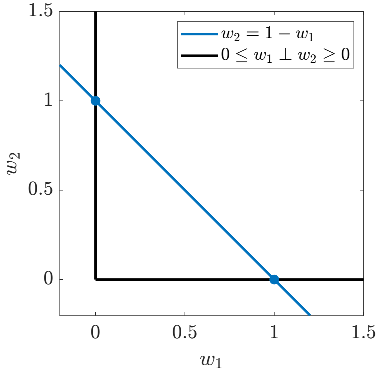
How complementarity constraints should **NOT** be used

Integrality conditions

$$w_1 \in \{0, 1\}$$

are equivalent to

$$\begin{aligned} w_2 &= 1 - w_1 \\ 0 \le w_1 &\perp w_2 \ge 0 \end{aligned}$$



- Feasible set consists of two isolate points.
- May converge only if initialized very close to solution.
- Problems with disjoint feasible regions should be avoided.
- Bad idea as $\sin(w\pi) = 0$ to obtain $w \in \mathbb{Z}$.

Biegler's guidelines for good MPEC modeling

Can be found in Chapter 11 of [Biegler, 2010].

1. To model discrete decisions, piecewise smooth functions, ..., start with a convex *lower level* problem, e.g.,

$$\min_{y} F(w_0)y \quad \text{s.t.} \quad y_\ell \le y \le y_u.$$

   Problem is parametric in switching function $F(w_0)$.

Biegler's guidelines for good MPEC modeling

Can be found in Chapter 11 of [Biegler, 2010].

1. To model discrete decisions, piecewise smooth functions, ..., start with a convex *lower level* problem, e.g.,

$$\min_{y} F(w_0)y \quad \text{s.t.} \quad y_\ell \leq y \leq y_u.$$

   Problem is parametric in switching function $F(w_0)$.

2. Whenever possible, formulate the problem such that constraints on upper level variables w_0 do not interfere with lower level constraints.
   If there are no constraints connecting w_0 and y, for $F(w_0) = 0$, $y \in [y_\ell, y_u]$ - avoiding disconnected feasible sets.

Biegler's guidelines for good MPEC modeling

Can be found in Chapter 11 of [Biegler, 2010].

1. To model discrete decisions, piecewise smooth functions, ..., start with a convex *lower level* problem, e.g.,

$$\min_{y} F(w_0)y \quad \text{s.t.} \quad y_\ell \leq y \leq y_u.$$

   Problem is parametric in switching function $F(w_0)$.

2. Whenever possible, formulate the problem such that constraints on upper level variables w_0 do not interfere with lower level constraints.
   If there are no constraints connecting w_0 and y, for $F(w_0) = 0$, $y \in [y_\ell, y_u]$ - avoiding disconnected feasible sets.

3. Apply KKT conditions, to reformulate MPEC to MPCC:

$$\begin{aligned} &F(w_0) - \lambda_\ell + \lambda_u = 0, \\ &0 \leq y - y_\ell \perp \lambda_\ell \geq 0, \\ &0 \leq y_u - y \perp \lambda_u \geq 0. \end{aligned}$$

Biegler's guidelines for good MPEC modeling

Can be found in Chapter 11 of [Biegler, 2010].

1. To model discrete decisions, piecewise smooth functions, ..., start with a convex *lower level* problem, e.g.,

$$\min_{y} F(w_0)y \quad \text{s.t.} \quad y_\ell \leq y \leq y_u.$$

   Problem is parametric in switching function $F(w_0)$.

2. Whenever possible, formulate the problem such that constraints on upper level variables w_0 do not interfere with lower level constraints.
   If there are no constraints connecting w_0 and y, for $F(w_0) = 0$, $y \in [y_\ell, y_u]$ - avoiding disconnected feasible sets.

3. Apply KKT conditions, to reformulate MPEC to MPCC:

$$\begin{aligned}
&F(w_0) - \lambda_\ell + \lambda_u = 0,\\
&0 \leq y - y_\ell \perp \lambda_\ell \geq 0,\\
&0 \leq y_u - y \perp \lambda_u \geq 0.
\end{aligned}$$

4. If possible, do some variable eliminations in the KKT conditions.

Modeling the $\mathrm{sign}(\cdot)$ and Heaviside step function

Set-valued Heaviside step function

$$\mathrm{step}(F(w)) = \begin{cases} \{1\}, & F(w) > 0, \\ [0,1], & F(w) = 0, \\ \{0\}, & F(w) < 0. \end{cases}$$

$$\mathrm{step}(F(w)) = \underset{y \in \mathbb{R}}{\mathrm{argmin}} \quad -F(w)y \tag{2a}$$

$$\text{s.t.} \quad 0 \leq y \leq 1. \tag{2b}$$

Modeling the $\mathrm{sign}(\cdot)$ and Heaviside step function

Set-valued Heaviside step function

$$\mathrm{step}(F(w)) = \begin{cases} \{1\}, & F(w) > 0, \\ [0,1], & F(w) = 0, \\ \{0\}, & F(w) < 0. \end{cases}$$

$$\mathrm{step}(F(w)) = \operatorname*{argmin}_{y \in \mathbb{R}} \quad -F(w)y \tag{2a}$$

$$\text{s.t.} \quad 0 \leq y \leq 1. \tag{2b}$$

KKT conditions of the LP (2):

$$\begin{aligned} &F(w) = \lambda_{\mathrm{p}} - \lambda_{\mathrm{n}}, \\ &0 \leq y \perp \lambda_{\mathrm{n}} \geq 0, \\ &0 \leq 1 - y \perp \lambda_{\mathrm{p}} \geq 0. \end{aligned}$$

Modeling the $\text{sign}(\cdot)$ and Heaviside step function

Set-valued Heaviside step function

$$\text{step}(F(w)) = \begin{cases} \{1\}, & F(w) > 0, \\ [0,1], & F(w) = 0, \\ \{0\}, & F(w) < 0. \end{cases}$$

$$\text{step}(F(w)) = \underset{y \in \mathbb{R}}{\arg\min} \quad -F(w)y \tag{2a}$$

$$\text{s.t.} \quad 0 \leq y \leq 1. \tag{2b}$$

KKT conditions of the LP (2):

$$\begin{aligned} &F(w) = \lambda_{\text{p}} - \lambda_{\text{n}}, \\ &0 \leq y \perp \lambda_{\text{n}} \geq 0, \\ &0 \leq 1 - y \perp \lambda_{\text{p}} \geq 0. \end{aligned}$$

- For $F(w) > 0 \implies y = 1$, $\lambda_{\text{n}} = 0$ positive part: $\lambda_{\text{p}} = \max(0, F(w)$.
- For $F(w) < 0 \implies y = 0$, $\lambda_{\text{p}} = 0$, negative part: $\lambda_{\text{n}} = \max(0, -F(w))$.
- $\text{sign}(F(w))$ by setting lower bound to -1 in LP (2), in this case $|F(w)| = F(w)y$.
- Cannot model discontinuous functions, only their set-valued extensions.

Modeling the $\mathrm{abs}(\cdot)$ function

Formulation 1:

- To model $z = |F(w)|$, use set-valued $y = \mathrm{sign}(F(w))$ function and identity:

$$z = |F(w)| = F(w)y.$$

Modeling the $\mathrm{abs}(\cdot)$ function

Formulation 1:

- To model $z = |F(w)|$, use set-valued $y = \mathrm{sign}(F(w))$ function and identity:

$$z = |F(w)| = F(w)y.$$

Formulation 2:

- To model $z = |F(w)|$, use set-valued $y = \mathrm{step}(F(w))$ function and identity:

$$z = |F(w)| = yF(w) + (1 - y)(-F(w)) = (2y - 1)F(w)$$

Modeling the $\mathrm{abs}(\cdot)$ function

Formulation 1:

- To model $z = |F(w)|$, use set-valued $y = \mathrm{sign}(F(w))$ function and identity:

$$z = |F(w)| = F(w)y.$$

Formulation 2:

- To model $z = |F(w)|$, use set-valued $y = \mathrm{step}(F(w))$ function and identity:

$$z = |F(w)| = yF(w) + (1-y)(-F(w)) = (2y-1)F(w)$$

Formulation 3:

- To model $z = |F(w)|$, use single complementarity constraints:

$$F(w) = \lambda^{\mathrm{p}} - \lambda^{\mathrm{n}}$$
$$0 \leq \lambda^{\mathrm{p}} \perp \lambda^{\mathrm{n}} \geq 0$$

- Combining the positive and negative part, $z = \lambda^{\mathrm{p}} + \lambda^{\mathrm{n}}$

Modeling the $\max(\cdot, \cdot)$ function

Formulation 1: to model $z = \max(F_1(w), F_2(w))$, use step LP with the argument $F_1(w) - F_2(w)$, i.e.:

$$\begin{aligned} \min_{y \in \mathbb{R}} \quad & (F_2(w) - F_1(w))y \\ \text{s.t.} \quad & 0 \leq y \leq 1. \end{aligned}$$

and compute $z = yF_1(w) + (1-y)F_2(w)$.

Modeling the $\max(\cdot,\cdot)$ function

Formulation 1: to model $z = \max(F_1(w), F_2(w))$, use step LP with the argument $F_1(w) - F_2(w)$, i.e.:

$$\begin{aligned} \min_{y \in \mathbb{R}} \quad & (F_2(w) - F_1(w))y \\ \text{s.t.} \quad & 0 \le y \le 1. \end{aligned}$$

and compute $z = yF_1(w) + (1-y)F_2(w)$.

Formulation 2: exploit that

$$z = \max(F_1(w), F_2(w)) = F_1(w) + \min(0, -(F_1(w) - F_2(w)))$$

and use that

Modeling the $\max(\cdot,\cdot)$ function

Formulation 1: to model $z = \max(F_1(w), F_2(w))$, use step LP with the argument $F_1(w) - F_2(w)$, i.e.:

$$\begin{aligned} \min_{y \in \mathbb{R}} \quad & (F_2(w) - F_1(w))y \\ \text{s.t.} \quad & 0 \leq y \leq 1. \end{aligned}$$

and compute $z = yF_1(w) + (1-y)F_2(w)$.

Formulation 2: exploit that

$$z = \max(F_1(w), F_2(w)) = F_1(w) + \min(0, -(F_1(w) - F_2(w)))$$

and use that

- $F_1(w) > F_2(w) \implies \min(0, -(F_1(w) - F_2(w))) = 0.$
- $F_1(w) < F_2(w) \implies \min(0, -(F_1(w) - F_2(w))) = -F_1(w) + F_2(w).$
- Using the positive and negative part of $F_1(w) - F_2(w)$ we have

$$\begin{aligned} & z = F_1(w) + \lambda_{\mathrm{n}}, \\ & F_1(w) - F_2(w) = \lambda_{\mathrm{p}} - \lambda_{\mathrm{n}}, \\ & 0 \leq \lambda_{\mathrm{p}} \perp \lambda_{\mathrm{n}} \geq 0. \end{aligned}$$

Modeling the $\min(\cdot,\cdot)$ function

Formulation 1: to model $z = \min(F_1(w), F_2(w))$, use step LP with the argument $F_1(w) - F_2(w)$, i.e.:

$$\begin{aligned} \min_{y \in \mathbb{R}} \quad & (F_2(w) - F_1(w))y \\ \text{s.t.} \quad & 0 \leq y \leq 1 \end{aligned}$$

and compute $z = (1-y)F_1(w) + yF_2(w)$.

Modeling the $\min(\cdot, \cdot)$ function

Formulation 1: to model $z = \min(F_1(w), F_2(w))$, use step LP with the argument $F_1(w) - F_2(w)$, i.e.:

$$\begin{aligned} \min_{y \in \mathbb{R}} \quad & (F_2(w) - F_1(w))y \\ \text{s.t.} \quad & 0 \leq y \leq 1 \end{aligned}$$

and compute $z = (1 - y)F_1(w) + yF_2(w)$.

Formulation 2: exploit that

$$z = \min(F_1(w), F_2(w)) = F_1(w) - \min(0, -(F_2(w) - F_1(w)))$$

and use that

Modeling the $\min(\cdot,\cdot)$ function

Formulation 1: to model $z = \min(F_1(w), F_2(w))$, use step LP with the argument $F_1(w) - F_2(w)$, i.e.:

$$\begin{aligned} \min_{y\in\mathbb{R}} \quad & (F_2(w) - F_1(w))y \\ \text{s.t.} \quad & 0 \le y \le 1 \end{aligned}$$

and compute $z = (1-y)F_1(w) + yF_2(w)$.

Formulation 2: exploit that

$$z = \min(F_1(w), F_2(w)) = F_1(w) - \min(0, -(F_2(w) - F_1(w)))$$

and use that

- $F_1(w) > F_2(w) \implies \min(0, -(F_2(w) - F_1(w))) = -F_1(w) + F_2(w)$
- $F_1(w) < F_2(w) \implies \min(0, -(F_2(w) - F_1(w))) = 0$
- Using positive and negative part of $F_1(w) - F_2(w)$ we have

$$\begin{aligned} z &= F_1(w) - \lambda_{\mathrm{n}} \\ F_2(w) - F_1(w) &= \lambda_{\mathrm{p}} - \lambda_{\mathrm{n}} \\ 0 \le \lambda_{\mathrm{p}} &\perp \lambda_{\mathrm{n}} \ge 0 \end{aligned}$$

Piecewise smooth functions and look-up tables

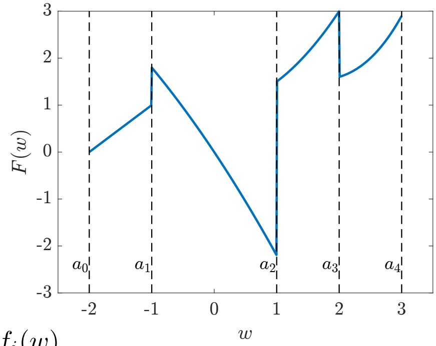
Introduced in [Raghunathan and Biegler, 2003]

Model a piecewise smooth functions $F : \mathbb{R} \to \mathbb{R}$ over intervals defined by grid points $a_0 < a_1 \ldots < a_n$

$$\min_{y \in \mathbb{R}^n} \quad \sum_{i=1}^{n} (w - a_i)(w - a_{i+1}) y_i$$

$$\text{s.t.} \quad \sum_{i=1}^{n} y_i = 1,$$

$$y \geq 0.$$



- Compute selectors variables y_i, and $z = \sum_{i=1}^{n} y_i f_i(w)$.
- If $w \in [a_{i-1}, a_i]$, then $(w - a_i)(w - a_{i+1}) \leq 0$, otherwise positive.
- Generalization to multi dimensional input and output spaces via Stewart's LP (Lecture 3), very efficient if Voronoi regions can be used.

Outline of the lecture

Vanishing (inequality) constraints

Studied in detail in [Achtziger and Kanzow, 2008]

Vanishing constraint

If $H_i(w) > 0$ then $G_i(w) \leq 0$, else if $H_i(w) = 0$ then $G_i(w) \in \mathbb{R}$ ($G_i(w) \leq 0$ vanishes)

Vanishing (inequality) constraints

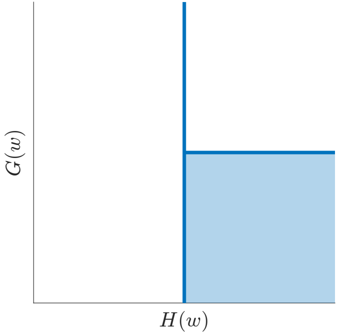
Studied in detail in [Achtziger and Kanzow, 2008]

Vanishing constraint

If $H_i(w) > 0$ then $G_i(w) \leq 0$, else if $H_i(w) = 0$ then $G_i(w) \in \mathbb{R}$ ($G_i(w) \leq 0$ vanishes)

- Logical formulation:

$$H(w) \geq 0,$$
$$H_i(w) > 0 \implies G_i(w) \leq 0, \quad i = 1, \ldots, m.$$



Vanishing (inequality) constraints

Studied in detail in [Achtziger and Kanzow, 2008]

Vanishing constraint

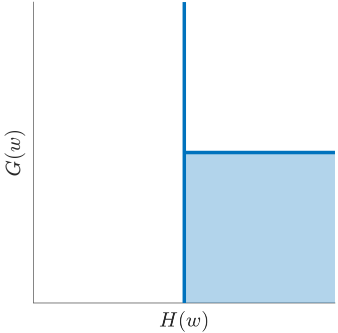
If $H_i(w) > 0$ then $G_i(w) \leq 0$, else if $H_i(w) = 0$ then $G_i(w) \in \mathbb{R}$ ($G_i(w) \leq 0$ vanishes)

- Logical formulation:

$$H(w) \geq 0,$$
$$H_i(w) > 0 \implies G_i(w) \leq 0, \quad i = 1, \dots, m.$$

- Nonlinear programming formulation:

$$H(w) \geq 0,$$
$$G_i(w) H_i(w) \leq 0, \quad i = 1, \dots, m.$$



Vanishing (inequality) constraints

Studied in detail in [Achtziger and Kanzow, 2008]

Vanishing constraint

If $H_i(w) > 0$ then $G_i(w) \leq 0$, else if $H_i(w) = 0$ then $G_i(w) \in \mathbb{R}$ ($G_i(w) \leq 0$ vanishes)

- Logical formulation:

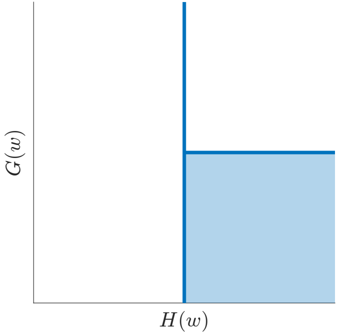
$$H(w) \geq 0,$$
$$H_i(w) > 0 \implies G_i(w) \leq 0, \quad i = 1, \ldots, m.$$

- Nonlinear programming formulation:

$$H(w) \geq 0,$$
$$G_i(w)H_i(w) \leq 0, \quad i = 1, \ldots, m.$$

- Lifted complementarity formulation:

$$G(w) - z \leq 0,$$
$$0 \leq z \perp H(w) \geq 0.$$



For $H_i(w) > 0$: $z = 0$, for $H_i(w) = 0$: $z = \max(G_i(w), 0)$.

State triggered constraints (vanishing equality constraints)

Introduced in [Szmuk et al., 2020]

State triggered constraints

If $H_i(w) < 0$ then $G_i(w) = 0$, otherwise if $H_i(w) \geq 0$ then $G_i(w) \in \mathbb{R}$

State triggered constraints (vanishing equality constraints)

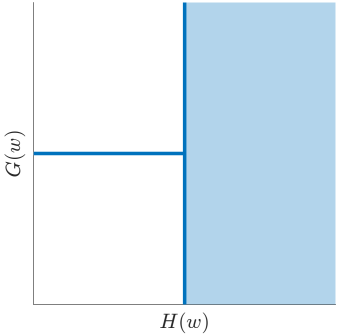
Introduced in [Szmuk et al., 2020]

State triggered constraints

If $H_i(w) < 0$ then $G_i(w) = 0$, otherwise if $H_i(w) \geq 0$ then $G_i(w) \in \mathbb{R}$

- Logical formulation:

$$H_i(w) < 0 \implies G_i(w) = 0, \quad i = 1, \ldots, m.$$



State triggered constraints (vanishing equality constraints)

Introduced in [Szmuk et al., 2020]

State triggered constraints

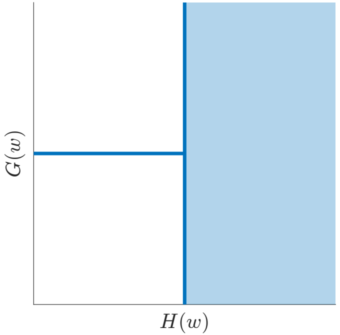
If $H_i(w) < 0$ then $G_i(w) = 0$, otherwise if $H_i(w) \geq 0$ then $G_i(w) \in \mathbb{R}$

- Logical formulation:

$$H_i(w) < 0 \implies G_i(w) = 0, \quad i = 1, \ldots, m.$$

- Complementarity formulation:

$$\begin{aligned} z \perp G(w), \\ 0 \leq H(w) + z \perp z \geq 0. \end{aligned}$$



State triggered constraints (vanishing equality constraints)

Introduced in [Szmuk et al., 2020]

State triggered constraints

If $H_i(w) < 0$ then $G_i(w) = 0$, otherwise if $H_i(w) \geq 0$ then $G_i(w) \in \mathbb{R}$

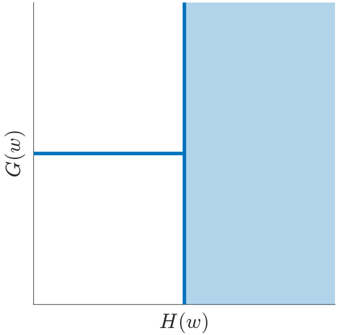
- Logical formulation:

$$H_i(w) < 0 \implies G_i(w) = 0, \quad i = 1, \dots, m.$$

- Complementarity formulation:

$$z \perp G(w),$$
$$0 \leq H(w) + z \perp z \geq 0.$$

- Interpretation:
  - $H_i(w) < 0 \implies z_i = -H_i(w) > 0 \implies G_i(w) = 0,$
  - $H_i(w) > 0 \implies z_i = 0 \implies G_i(w) \in \mathbb{R},$
  - $H_i(w) = 0 \implies z_i = 0 \implies G_i(w) \in \mathbb{R}.$



Sparsity optimization problems

Studied in [Feng et al., 2018, Kanzow et al., 2024]

- $\|w\|_0$, the ℓ_0 "norm" of $w \in \mathbb{R}^n$, is the number of nonzero elements in this vector.
- In practice, relaxed via ℓ_1.
- Complementarity constraints allow exact reformulations (same set of minimizers).

Sparsity optimization (SPO)

$$\begin{aligned} \min_{w \in \mathbb{R}^n} \quad & f(w) + \rho \|w\|_0 \\ \text{s.t.} \quad & g(w) = 0, \\ & h(w) \geq 0. \end{aligned}$$

$$e = [1, \dots, 1]^\top$$

Sparsity optimization problems

Studied in [Feng et al., 2018, Kanzow et al., 2024]

- $\|w\|_0$, the ℓ_0 "norm" of $w \in \mathbb{R}^n$, is the number of nonzero elements in this vector.
- In practice, relaxed via ℓ_1.
- Complementarity constraints allow exact reformulations (same set of minimizers).

Sparsity optimization (SPO)

$$\begin{aligned} \min_{w \in \mathbb{R}^n} \quad & f(w) + \rho\|w\|_0 \\ \text{s.t.} \quad & g(w) = 0, \\ & h(w) \geq 0. \end{aligned}$$

Linear SPO [Feng et al., 2018]

$$\begin{aligned} \min_{w,z \in \mathbb{R}^n} \quad & f(w) + \rho(n - e^\top z) \\ \text{s.t.} \quad & g(w) = 0, \\ & h(w) \geq 0, \\ & 0 \leq z \leq e \\ & w \perp z \end{aligned}$$

$e = [1, \dots, 1]^\top$

Sparsity optimization problems

Studied in [Feng et al., 2018, Kanzow et al., 2024]

- $\|w\|_0$, the ℓ_0 "norm" of $w \in \mathbb{R}^n$, is the number of nonzero elements in this vector.
- In practice, relaxed via ℓ_1.
- Complementarity constraints allow exact reformulations (same set of minimizers).

Sparsity optimization (SPO)

$$\begin{aligned} \min_{w \in \mathbb{R}^n} \quad & f(w) + \rho \|w\|_0 \\ \text{s.t.} \quad & g(w) = 0, \\ & h(w) \geq 0. \end{aligned}$$

Linear SPO [Feng et al., 2018]

$$\begin{aligned} \min_{w,z \in \mathbb{R}^n} \quad & f(w) + \rho(n - e^\top z) \\ \text{s.t.} \quad & g(w) = 0, \\ & h(w) \geq 0, \\ & 0 \leq z \leq e \\ & w \perp z \end{aligned}$$

Quadratic SPO [Kanzow et al., 2024]

$$\begin{aligned} \min_{w \in \mathbb{R}^n} \quad & f(w) + \frac{\rho}{2} \sum_i^1 z_i(z_i - 2) \\ \text{s.t.} \quad & g(w) = 0, \\ & h(w) \geq 0, \\ & w \perp z \end{aligned}$$

$$e = [1, \ldots, 1]^\top$$

Sparsity optimization with full complementarity constraints

Derived in [Feng et al., 2018]

Sparsity optimization (SPO)

$$\begin{aligned} \min_{w \in \mathbb{R}^n} \quad & f(w) + \rho \|w\|_0 \\ \text{s.t.} \quad & g(w) = 0, \\ & h(w) \geq 0. \end{aligned}$$

Full complementarity SPO [Feng et al., 2018]

$$\begin{aligned} \min_{w, w^{\pm}, z \in \mathbb{R}^n} \quad & f(w) + \rho \sum_{i=1}^{n} (1 - z_i) \\ \text{s.t.} \quad & g(w) = 0, \\ & w = w^+ - w^-, \\ & 0 \leq z \perp w^+ + w^- \geq 0, \\ & 0 \leq w^+ \perp w^- \geq 0, \\ & z \leq e. \end{aligned}$$

- It can be deduced that at optimal solutions $z_i = \begin{cases} 0, & w_i \neq 0 \\ 1, & w_i = 0 \end{cases}$
- The same reformulations can be used to handle cardinality constraints $\|w\|_0 \leq \kappa$, $\kappa \in \mathbb{N}$, but may have additional local minima [Kanzow et al., 2024].

Calculus of logical operations

Logic with complementarity from [Pozharskiy et al., 2024], collection of logical relations [ApS, 2024, Chapter 9]

Logical operators for $x, y \geq 0$

$$
\begin{aligned}
& z > 0 \quad \text{(true)}, \qquad z = 0 \quad \text{(false)} \\
& z = x \vee y \iff z \geq x,\ z \geq y,\ z \leq x + y, \\
& z = x \wedge y \iff z \leq x,\ z \leq y,\ z \geq x + y - \max(x, y), \\
& x \implies z \iff x \leq z
\end{aligned}
$$

max via convex optimization

$$
\max(x, y) = \min_{z \in \mathbb{R}} z \quad \text{s.t. } z \geq x,\ z \geq y
$$

- If $z = 0$ (false) then $g(w) = 0$:

$$
-zM \leq g(w) \leq zM, \quad M \gg 1.
$$

- If $z = 0$ (false) then $h(w) \leq 0$:

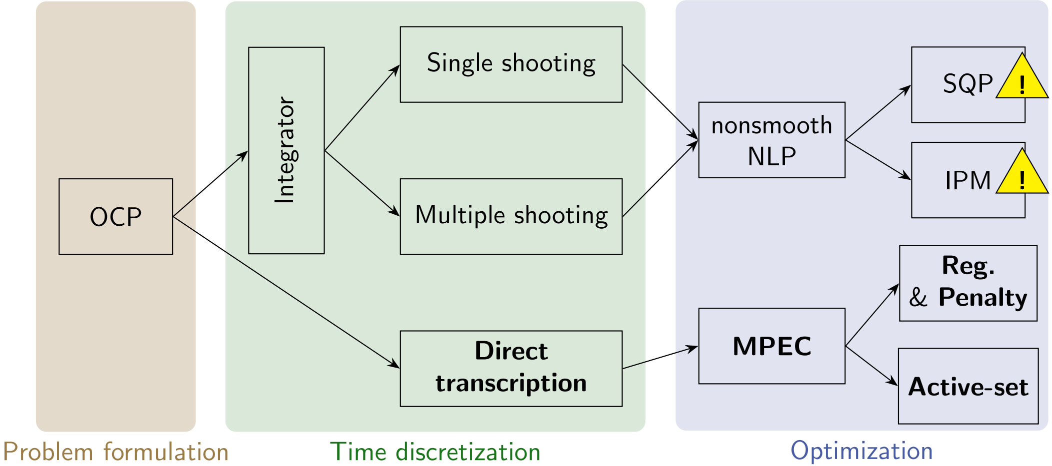
$$
h(w) \leq zM, \quad M \gg 1.
$$

- If $z > 0$ then $w = 1$:

$$
w = \mathrm{step}(z) \quad (\text{otherwise } w \in [0, 1])
$$

Work flow in nonsmooth direct optimal control

First discretize, then optimize.



OCP = Optimal Control Problem

NLP = Nonlinear Program

MPEC = Mathematical Program with Equilibrium Constraints

SQP = Sequential Quadratic Programming

IPM = Interior-Point Method

Reg. = Regularization

Outline of the lecture

Refresher on the Karush-Kuhn-Tucker (KKT) conditions

Nonlinear Program (NLP)

$$
\begin{aligned}
\min_{w \in \mathbb{R}^n} \; & f(w) \\
\text{s.t. } & g(w) = 0, \\
& h(w) \geq 0.
\end{aligned}
$$

$$\mathcal{L}(w, \lambda, \mu) = f(w) - \lambda^\top g(w) - \mu^\top h(w)$$

Refresher on the Karush-Kuhn-Tucker (KKT) conditions

Nonlinear Program (NLP)

$$\begin{aligned} \min_{w \in \mathbb{R}^n} \; & f(w) \\ \text{s.t. } & g(w) = 0, \\ & h(w) \geq 0. \end{aligned}$$

$$\mathcal{L}(w, \lambda, \mu) = f(w) - \lambda^\top g(w) - \mu^\top h(w)$$

Definition (LICQ)

A point w satisfies LICQ if

$$[\nabla g(w), \quad \nabla h_{\mathcal{A}}(w)]$$

is full column rank.

Active set $\mathcal{A} = \{i \mid h_i(w) = 0\}$

Refresher on the Karush-Kuhn-Tucker (KKT) conditions

Nonlinear Program (NLP)

$$\min_{w \in \mathbb{R}^n} \; f(w)$$
$$\text{s.t. } g(w) = 0,$$
$$h(w) \geq 0.$$

$$\mathcal{L}(w, \lambda, \mu) = f(w) - \lambda^\top g(w) - \mu^\top h(w)$$

Definition (LICQ)

A point w satisfies LICQ if

$$[\nabla g(w), \quad \nabla h_{\mathcal{A}}(w)]$$

is full column rank.

Active set $\mathcal{A} = \{i \mid h_i(w) = 0\}$

Theorem (KKT conditions - FONC for constrained optimization)

Let f, g, h be $\mathcal{C}^1$. If w^* is a (local) minimizer ***and*** *satisfies LICQ, then there are unique vectors λ^* and μ^* such that (w^*, λ^*, μ^*) satisfies:*

$$\nabla_w \mathcal{L}(w^*, \mu^*, \lambda^*) = 0, \quad \mu^* \geq 0, \qquad \textit{dual feasibility}$$
$$g(w^*) = 0, \quad h(w^*) \geq 0 \qquad \textit{primal feasibility}$$
$$\mu_i^* h_i(w^*) = 0, \quad \forall\, i \qquad \textit{complementary slackness}$$

Constraint qualifications: LICQ and MFCQ

Definition (LICQ)

A point w satisfies LICQ if

$$[\nabla g\,(w)\,, \quad \nabla h_{\mathcal{A}}\,(w)]$$

is full column rank.

Active set $\mathcal{A} = \{i \mid h_i(w) = 0\}$

Constraint qualifications: LICQ and MFCQ

Definition (LICQ)

A point w satisfies LICQ if

$$[\nabla g(w), \quad \nabla h_{\mathcal{A}}(w)]$$

is full column rank.

Active set $\mathcal{A} = \{i \mid h_i(w) = 0\}$

Definition (MFCQ)

A point w satisfies the Mangasarian–Fromovitz constraint qualification (MFCQ), if $\nabla g(w)$ has full column rank, and if there exist a direction $d \in \mathbb{R}^n$ such that:

$$\nabla g(w)^\top d = 0, \quad \nabla h_i(w)^\top d > 0.$$

Constraint qualifications: LICQ and MFCQ

Definition (LICQ)

A point w satisfies LICQ if

$$[\nabla g\left(w\right), \quad \nabla h_{\mathcal{A}}\left(w\right)]$$

is full column rank.

Active set $\mathcal{A} = \{i \mid h_i(w) = 0\}$

- LICQ implies MFCQ. Direction can be found by solving

$$\nabla g(w)^\top d = 0, \quad \nabla h_i(w)^\top d = e.$$

- Both require full rank of $\nabla g\left(w\right)$.
- Example MFCQ holds, LICQ does not: $w_2 - w_1^2 \geq 0$ and $w_2 - w_1^4 \geq 0$ at $w = 0$.

Definition (MFCQ)

A point w satisfies the Mangasarian–Fromovitz constraint qualification (MFCQ), if $\nabla g\left(w\right)$ has full column rank, and if there exist a direction $d \in \mathbb{R}^n$ such that:

$$\nabla g(w)^\top d = 0, \quad \nabla h_i(w)^\top d > 0.$$

Constraint qualifications: LICQ and MFCQ

Definition (LICQ)

A point w satisfies LICQ if

$$[\nabla g\left(w\right), \quad \nabla h_{\mathcal{A}}\left(w\right)]$$

is full column rank.

Active set $\mathcal{A} = \{i \mid h_i(w) = 0\}$

- LICQ implies MFCQ. Direction can be found by solving

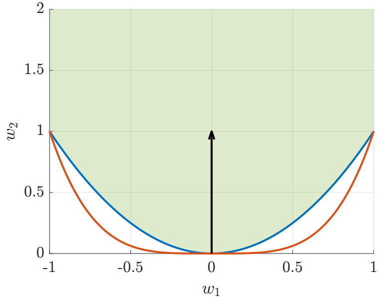
$$\nabla g(w)^\top d = 0, \quad \nabla h_i(w)^\top d = e.$$

- Both require full rank of $\nabla g\left(w\right)$.
- Example MFCQ holds, LICQ does not: $w_2 - w_1^2 \geq 0$ and $w_2 - w_1^4 \geq 0$ at $w = 0$.

Definition (MFCQ)

A point w satisfies the Mangasarian–Fromovitz constraint qualification (MFCQ), if $\nabla g\left(w\right)$ has full column rank, and if there exist a direction $d \in \mathbb{R}^n$ such that:

$$\nabla g(w)^\top d = 0, \quad \nabla h_i(w)^\top d > 0.$$



Applying KKT conditions to MPECs

Example from [Scheel and Scholtes, 2000].

Example linear MPCC

$$
\begin{aligned}
\min_{w \in \mathbb{R}^3} \quad & w_1 + w_2 - w_3 \\
\text{s.t.} \quad & h_1(w) = 4w_1 - w_3 \geq 0, && \mid \mu_1, \\
& h_2(w) = 4w_2 - w_3 \geq 0, && \mid \mu_2, \\
& w_1 \geq 0, && \mid \mu_G, \\
& w_2 \geq 0, && \mid \mu_H, \\
& w_1 w_2 \leq 0, && \mid \mu_{GH}.
\end{aligned}
$$

Applying KKT conditions to MPECs

Example from [Scheel and Scholtes, 2000].

Example linear MPCC

$$
\begin{aligned}
\min_{w \in \mathbb{R}^3} \quad & w_1 + w_2 - w_3 \\
\text{s.t.} \quad & h_1(w) = 4w_1 - w_3 \geq 0, && \mid \mu_1, \\
& h_2(w) = 4w_2 - w_3 \geq 0, && \mid \mu_2, \\
& w_1 \geq 0, && \mid \mu_G, \\
& w_2 \geq 0, && \mid \mu_H, \\
& w_1 w_2 \leq 0, && \mid \mu_{GH}.
\end{aligned}
$$

- Global optimum $w^* = (0, 0, 0)$.
- All constraints are active at the solution.
- Use necessary KKT conditions to compute the optimal Lagrange multipliers.

Global optimum may not be a KKT point

Example from [Scheel and Scholtes, 2000].

KKT conditions applied to a MPCC

$$0 = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} - \mu_1^* \begin{bmatrix} 4 \\ 0 \\ -1 \end{bmatrix} - \mu_2^* \begin{bmatrix} 0 \\ 4 \\ -1 \end{bmatrix} - \mu_G^* \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - \mu_H^* \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + \mu_{GH}^* \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

From the non-negativity of the multipliers and this condition we obtain an inconsistent system:

$$\begin{aligned} &\mu_G^* = 1 - 4\mu_1^*,\ \mu_G^*, \mu_1^* \geq 0 \implies \mu_1^* \in [0, 0.25], \\ &\mu_H^* = 1 - 4\mu_2^*,\ \mu_H^*, \mu_2^* \geq 0 \implies \mu_2^* \in [0, 0.25], \\ &\mu_1^* + \mu_2^* = 1. \end{aligned}$$

Global optimum may not be a KKT point

Example from [Scheel and Scholtes, 2000].

KKT conditions applied to a MPCC

$$0 = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} - \mu_1^* \begin{bmatrix} 4 \\ 0 \\ -1 \end{bmatrix} - \mu_2^* \begin{bmatrix} 0 \\ 4 \\ -1 \end{bmatrix} - \mu_G^* \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - \mu_H^* \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + \mu_{GH}^* \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

From the non-negativity of the multipliers and this condition we obtain an inconsistent system:

$$\begin{aligned} &\mu_G^* = 1 - 4\mu_1^*, \ \mu_G^*, \mu_1^* \geq 0 \implies \mu_1^* \in [0, 0.25], \\ &\mu_H^* = 1 - 4\mu_2^*, \ \mu_H^*, \mu_2^* \geq 0 \implies \mu_2^* \in [0, 0.25], \\ &\mu_1^* + \mu_2^* = 1. \end{aligned}$$

Global optimum not a KKT point.

Reason: constraint qualifications violated.

Outline of MPEC optimality theory

- LICQ and MFCQ violated at all feasible points of an NLP formulation of an MPEC $w_1 \geq 0, w_2 \geq 0, w_1 w_2 \leq 0$.
- KKT conditions need constraint qualifications to hold.
- We need better first-order necessary optimality conditions, that hold under weaker, MPEC-tailored constraint qualifications.
- We show primal (no multipliers) and dual (with Lagrange multipliers) optimality conditions for MPECs, and when they are equal.

Outline of MPEC optimality theory

- LICQ and MFCQ violated at all feasible points of an NLP formulation of an MPEC $w_1 \geq 0, w_2 \geq 0, w_1 w_2 \leq 0$.
- KKT conditions need constraint qualifications to hold.
- We need better first-order necessary optimality conditions, that hold under weaker, MPEC-tailored constraint qualifications.
- We show primal (no multipliers) and dual (with Lagrange multipliers) optimality conditions for MPECs, and when they are equal.
- Having an understanding of optimality conditions is essential for designing efficient numerical methods.

Tangent cone definition

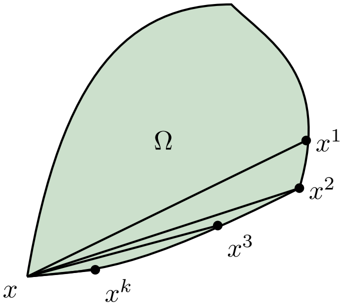
Nonlinear program (NLP)

$$\min_{w \in \mathbb{R}^n} \quad f(w) \tag{3a}$$

$$\text{s.t.} \quad g(w) = 0, \tag{3b}$$

$$h(w) \geq 0, \tag{3c}$$

$$\Omega = \{w \in \mathbb{R}^n \mid g(w) = 0, h(w) \geq 0\}$$



Definition (Bouligand tangent cone[1])

The (Bouligand) tangent cone at $w \in \Omega$ is defined as the set:

$$\mathcal{T}_\Omega(w) = \{d \in \mathbb{R}^n \mid \exists \{w^k\} \subset \Omega, \{t^k\} \subset \mathbb{R}_{\geq 0} : \lim_{k \to \infty} t^k = 0, \lim_{k \to \infty} w^k = w, \lim_{k \to \infty} \frac{w^k - w}{t^k} = d\}$$

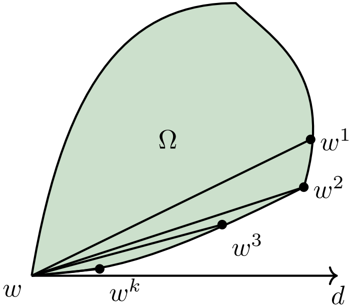
[1] Georges Louis Bouligand (1889 – 1979), a French mathematician, introduced this tangent cone definition in 1932.

Tangent cone definition

Nonlinear program (NLP)

$$\min_{w \in \mathbb{R}^n} \quad f(w) \tag{3a}$$
$$\text{s.t.} \quad g(w) = 0, \tag{3b}$$
$$h(w) \geq 0, \tag{3c}$$

$$\Omega = \{w \in \mathbb{R}^n \mid g(w) = 0, h(w) \geq 0\}$$



Definition (Bouligand tangent cone[1])

The (Bouligand) tangent cone at $w \in \Omega$ is defined as the set:

$$\mathcal{T}_\Omega(w) = \{d \in \mathbb{R}^n \mid \exists \{w^k\} \subset \Omega, \{t^k\} \subset \mathbb{R}_{\geq 0} : \lim_{k \to \infty} t^k = 0, \lim_{k \to \infty} w^k = w, \lim_{k \to \infty} \frac{w^k - w}{t^k} = d\}$$

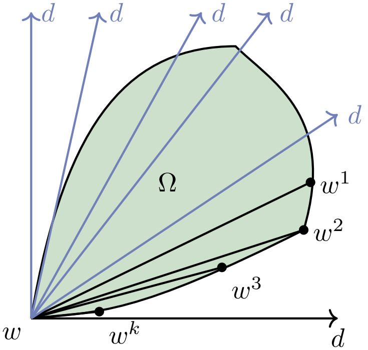
[1] Georges Louis Bouligand (1889 – 1979), a French mathematician, introduced this tangent cone definition in 1932.

Tangent cone definition

Nonlinear program (NLP)

$$\min_{w \in \mathbb{R}^n} \quad f(w) \tag{3a}$$
$$\text{s.t.} \quad g(w) = 0, \tag{3b}$$
$$h(w) \geq 0, \tag{3c}$$

$$\Omega = \{w \in \mathbb{R}^n \mid g(w) = 0, h(w) \geq 0\}$$



Definition (Bouligand tangent cone[1])

The (Bouligand) tangent cone at $w \in \Omega$ is defined as the set:

$$\mathcal{T}_\Omega(w) = \{d \in \mathbb{R}^n \mid \exists \{w^k\} \subset \Omega, \{t^k\} \subset \mathbb{R}_{\geq 0} : \lim_{k\to\infty} t^k = 0, \lim_{k\to\infty} w^k = w, \lim_{k\to\infty} \frac{w^k - w}{t^k} = d\}$$

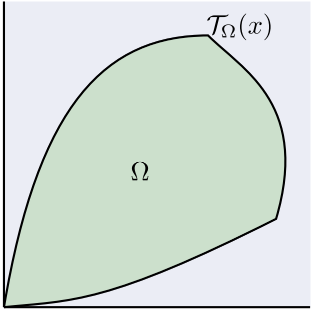
[1] Georges Louis Bouligand (1889 – 1979), a French mathematician, introduced this tangent cone definition in 1932.

Tangent cone definition

Nonlinear program (NLP)

$$\min_{w\in\mathbb{R}^n} \quad f(w) \tag{3a}$$
$$\text{s.t.} \quad g(w) = 0, \tag{3b}$$
$$h(w) \geq 0, \tag{3c}$$

$$\Omega = \{w \in \mathbb{R}^n \mid g(w) = 0, h(w) \geq 0\}$$



Definition (Bouligand tangent cone[1])

The (Bouligand) tangent cone at $w \in \Omega$ is defined as the set:

$$\mathcal{T}_\Omega(w) = \{d \in \mathbb{R}^n \mid \exists \{w^k\} \subset \Omega, \{t^k\} \subset \mathbb{R}_{\geq 0} : \lim_{k\to\infty} t^k = 0, \lim_{k\to\infty} w^k = w, \lim_{k\to\infty} \frac{w^k - w}{t^k} = d\}$$

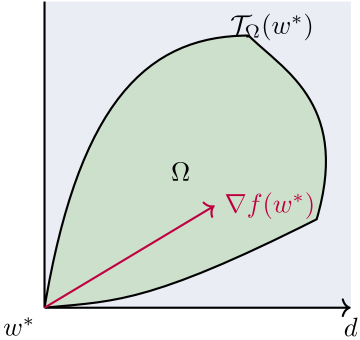
[1] Georges Louis Bouligand (1889 – 1979), a French mathematician, introduced this tangent cone definition in 1932.

First-order necessary conditions for optimality

Nonlinear program (NLP)

$$\begin{aligned}\min_{w\in\mathbb{R}^n} \quad & f(w)\\ \text{s.t.} \quad & g(w)=0,\\ & h(w)\geq 0,\end{aligned}$$

$$\Omega = \{w \in \mathbb{R}^n \mid g(w)=0, h(w)\geq 0\}$$



Theorem (First-Order Necessary Conditions)

If $w^* \in \Omega$ is a local minimum of the NLP (3) then it holds that

$$\nabla f(w^*)^\top d \geq 0 \text{ for all tangents } d \in \mathcal{T}_\Omega(w^*). \qquad (4)$$

The definition of is B-stationarity

Definition (B-stationarity)

A point $w^* \in \Omega$ satisfying (4) is called a Bouligand stationary (B-stationary) point.

Definition (Linearized feasible cone)

Let $w \in \Omega$ and $\mathcal{A}(w) = \{i \mid h_i(w) = 0\}$. The linearized feasible cone is defined as the set:

$$\mathcal{T}_{\Omega}^{\mathrm{lin}}(w) = \{d \in \mathbb{R}^n \mid \nabla g(w)^\top d = 0, \nabla h_i(w)^\top d \geq 0, i \in \mathcal{A}(w)\}$$

The definition of is B-stationarity

Definition (B-stationarity)

A point $w^* \in \Omega$ satisfying (4) is called a Bouligand stationary (B-stationary) point.

Definition (Linearized feasible cone)

Let $w \in \Omega$ and $\mathcal{A}(w) = \{i \mid h_i(w) = 0\}$. The linearized feasible cone is defined as the set:

$$\mathcal{T}_{\Omega}^{\mathrm{lin}}(w) = \{d \in \mathbb{R}^n \mid \nabla g(w)^\top d = 0, \nabla h_i(w)^\top d \geq 0, i \in \mathcal{A}(w)\}$$

Standard nonlinear programming approach:

1. If a constraint qualification holds (e.g. LICQ), $\mathcal{T}_{\Omega}(w) = \mathcal{T}_{\Omega}^{\mathrm{lin}}(w)$.

The definition of is B-stationarity

Definition (B-stationarity)

A point $w^* \in \Omega$ satisfying (4) is called a Bouligand stationary (B-stationary) point.

Definition (Linearized feasible cone)

Let $w \in \Omega$ and $\mathcal{A}(w) = \{i \mid h_i(w) = 0\}$. The linearized feasible cone is defined as the set:

$$\mathcal{T}_\Omega^{\text{lin}}(w) = \{d \in \mathbb{R}^n \mid \nabla g(w)^\top d = 0, \nabla h_i(w)^\top d \geq 0, i \in \mathcal{A}(w)\}$$

Standard nonlinear programming approach:

1. If a constraint qualification holds (e.g. LICQ), $\mathcal{T}_\Omega(w) = \mathcal{T}_\Omega^{\text{lin}}(w)$.
2. Use Farkas' lemma, obtain from Theorem 6 the Karush–Kuhn–Tucker (KKT) conditions.

The definition of is B-stationarity

Definition (B-stationarity)

A point $w^* \in \Omega$ satisfying (4) is called a Bouligand stationary (B-stationary) point.

Definition (Linearized feasible cone)

Let $w \in \Omega$ and $\mathcal{A}(w) = \{i \mid h_i(w) = 0\}$. The linearized feasible cone is defined as the set:

$$\mathcal{T}_{\Omega}^{\mathrm{lin}}(w) = \{d \in \mathbb{R}^n \mid \nabla g(w)^\top d = 0, \nabla h_i(w)^\top d \geq 0, i \in \mathcal{A}(w)\}$$

Standard nonlinear programming approach:

1. If a constraint qualification holds (e.g. LICQ), $\mathcal{T}_{\Omega}(w) = \mathcal{T}_{\Omega}^{\mathrm{lin}}(w)$.
2. Use Farkas' lemma, obtain from Theorem 6 the Karush–Kuhn–Tucker (KKT) conditions.
3. Solve a sequence of approximations of the KKT conditions to find a solution candidate (e.g., with `IPOPT` [Wächter and Biegler, 2006]).

MPEC

$$\min_{w \in \mathbb{R}^n} \quad f(w) \tag{5a}$$

$$\text{s.t.} \quad g(w) = 0, \tag{5b}$$

$$h(w) \geq 0, \tag{5c}$$

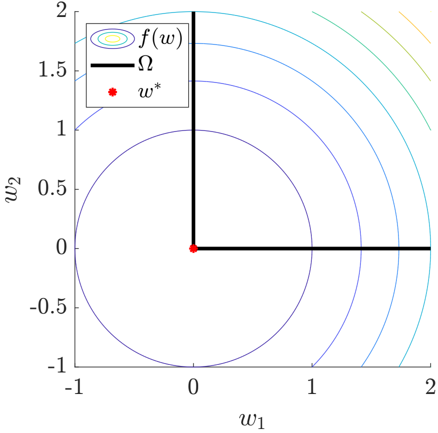
$$0 \leq w_1 \perp w_2 \geq 0, \tag{5d}$$

$$w = (w_0, w_1, w_2) \in \mathbb{R}^n,\ \ w_0 \in \mathbb{R}^p,\ \ w_1, w_2 \in \mathbb{R}^m,$$

$$\Omega = \{x \in \mathbb{R}^n \mid g(w) = 0, h(w) \geq 0, 0 \leq w_1 \perp w_2 \geq 0\},$$

MPEC active sets:

$$\mathcal{I}_{+0}(w) = \{i \in \{1, \dots, m\} \mid w_{1,i} > 0, w_{2,i} = 0\},$$
$$\mathcal{I}_{0+}(w) = \{i \in \{1, \dots, m\} \mid w_{1,i} = 0, w_{2,i} > 0\},$$
$$\mathcal{I}_{00}(w) = \{i \in \{1, \dots, m\} \mid w_{1,i} = 0, w_{2,i} = 0\}.$$



MPEC

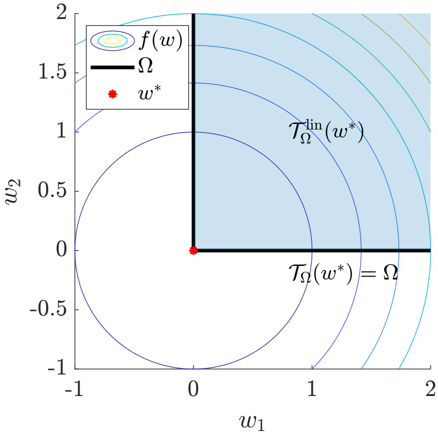
$$\min_{w \in \mathbb{R}^n} \quad f(w) \tag{5a}$$
$$\text{s.t.} \quad g(w) = 0, \tag{5b}$$
$$h(w) \geq 0, \tag{5c}$$
$$0 \leq w_1 \perp w_2 \geq 0, \tag{5d}$$

$$w = (w_0, w_1, w_2) \in \mathbb{R}^n, \ w_0 \in \mathbb{R}^p, \ w_1, w_2 \in \mathbb{R}^m,$$

$$\Omega = \{x \in \mathbb{R}^n \mid g(w) = 0, h(w) \geq 0, 0 \leq w_1 \perp w_2 \geq 0\},$$

MPEC active sets:

$$\mathcal{I}_{+0}(w) = \{i \in \{1, \ldots, m\} \mid w_{1,i} > 0, w_{2,i} = 0\},$$
$$\mathcal{I}_{0+}(w) = \{i \in \{1, \ldots, m\} \mid w_{1,i} = 0, w_{2,i} > 0\},$$
$$\mathcal{I}_{00}(w) = \{i \in \{1, \ldots, m\} \mid w_{1,i} = 0, w_{2,i} = 0\}.$$



In most interesting cases: $\mathcal{T}_{\Omega}(w) \neq \mathcal{T}_{\Omega}^{\text{lin}}(w)$.

MPEC-tailored linearized feasible cone

Definition (MPEC-linearized feasible cone [Flegel and Kanzow, 2005])

Let $w \in \Omega$ and $d = (d_0, d_1, d_2) \in \mathbb{R}^{p+2m}$. The MPEC-linearized feasible cone is the set

$$\begin{aligned}
\mathcal{T}_\Omega^{\mathrm{MPEC}}(w) = \{d \in \mathbb{R}^n \mid & \nabla g(w)^\top d = 0, \\
& \nabla h_i(w)^\top d \geq 0, \forall i \in \mathcal{A}(w), \\
& d_{1,i} = 0, \forall i \in \mathcal{I}_{0+}(w), \\
& d_{2,i} = 0, \forall i \in \mathcal{I}_{+0}(w), \\
& 0 \leq d_{1,i} \perp d_{2,i} \geq 0, \forall i \in \mathcal{I}_{00}(w)\}.
\end{aligned}$$

MPEC-tailored linearized feasible cone

Definition (MPEC-linearized feasible cone [Flegel and Kanzow, 2005])

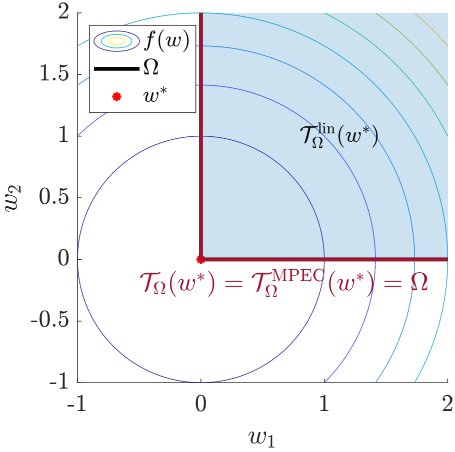
Let $w \in \Omega$ and $d = (d_0, d_1, d_2) \in \mathbb{R}^{p+2m}$. The MPEC-linearized feasible cone is the set

$$\begin{aligned}\mathcal{T}_{\Omega}^{\mathrm{MPEC}}(w) = \{d \in \mathbb{R}^n \mid \nabla g(w)^\top d = 0,\\ \nabla h_i(w)^\top d \geq 0, \forall i \in \mathcal{A}(w),\\ d_{1,i} = 0, \forall i \in \mathcal{I}_{0+}(w),\\ d_{2,i} = 0, \forall i \in \mathcal{I}_{+0}(w),\\ 0 \leq d_{1,i} \perp d_{2,i} \geq 0, \forall i \in \mathcal{I}_{00}(w)\}.\end{aligned}$$

Lemma (Lemma 3.2 in [Flegel and Kanzow, 2005])

Let $w \in \Omega$, then for the MPEC (5) it holds that:

$$\mathcal{T}_{\Omega}(w) \subseteq \mathcal{T}_{\Omega}^{\mathrm{MPEC}}(w) \subseteq \mathcal{T}_{\Omega}^{\mathrm{lin}}(w).$$



Primal stationarity conditions for MPECs

First order necessary optimality conditions - revisited

Theorem (Theorem 6 extended, [Luo et al., 1996])

Let $\mathcal{T}_\Omega(w) = \mathcal{T}_\Omega^{\mathrm{MPEC}}(w)$, and $w^* \in \Omega$ be a local minimizer of the MPEC (5), then it holds that

$$\nabla f(w^*)^\top d \geq 0 \ \ \textit{for all } d \in \mathcal{T}_\Omega^{\mathrm{MPEC}}(w^*), \tag{6}$$

or equivalent to (6), $d = 0$ is a local minimizer of the following optimization problem:

$$\min_{d \in \mathbb{R}^n} \quad \nabla f(w^*)^\top d \quad \text{s.t.} \quad d \in \mathcal{T}_\Omega^{\mathrm{MPEC}}(w^*). \tag{7}$$

- In interesting cases: $\mathcal{T}_\Omega(w) = \mathcal{T}_\Omega^{\mathrm{MPEC}}(w) \neq \mathcal{T}_\Omega^{\mathrm{lin}}(w)$.

Primal stationarity conditions for MPECs

First order necessary optimality conditions - revisited

Theorem (Theorem 6 extended, [Luo et al., 1996])

Let $\mathcal{T}_\Omega(w) = \mathcal{T}_\Omega^{\mathrm{MPEC}}(w)$, and $w^* \in \Omega$ be a local minimizer of the MPEC (5), then it holds that

$$\nabla f(w^*)^\top d \geq 0 \ \ \textit{for all } d \in \mathcal{T}_\Omega^{\mathrm{MPEC}}(w^*), \tag{6}$$

or equivalent to (6), $d = 0$ is a local minimizer of the following optimization problem:

$$\min_{d \in \mathbb{R}^n} \quad \nabla f(w^*)^\top d \quad \text{s.t.} \quad d \in \mathcal{T}_\Omega^{\mathrm{MPEC}}(w^*). \tag{7}$$

- In interesting cases: $\mathcal{T}_\Omega(w) = \mathcal{T}_\Omega^{\mathrm{MPEC}}(w) \neq \mathcal{T}_\Omega^{\mathrm{lin}}(w)$.
- Use the linear program with complementarity constraints. (LPEC) (7) instead of the KKT conditions.
- If $\mathcal{I}_{00}(w) \neq \emptyset$, LPEC is nonconvex, problem combinatorial in nature.
- If $\mathcal{I}_{00}(w) = \emptyset$, LPEC reduces to LP, at $d = 0$ LP KKT conditions $=$ NLP KKT conditions.

Pieces of the MPEC: the Tight Nonlinear Program (TNLP)

Regular NLPs, used to define MPEC-specific concepts.

MPEC active sets

$$\begin{aligned}
\mathcal{I}_{+0}(w) &= \{i \in \{1,\dots,m\} \mid w_{1,i} > 0, w_{2,i} = 0\},\\
\mathcal{I}_{0+}(w) &= \{i \in \{1,\dots,m\} \mid w_{1,i} = 0, w_{2,i} > 0\},\\
\mathcal{I}_{00}(w) &= \{i \in \{1,\dots,m\} \mid w_{1,i} = 0, w_{2,i} = 0\}.
\end{aligned}$$

Pieces of the MPEC: the Tight Nonlinear Program (TNLP)

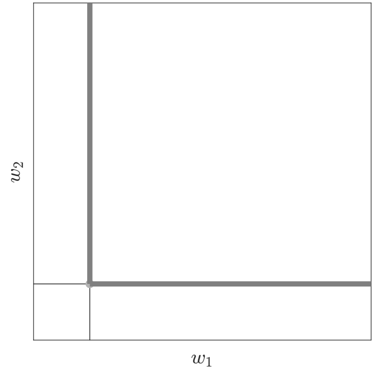
Regular NLPs, used to define MPEC-specific concepts.

MPEC active sets

$$\mathcal{I}_{+0}(w) = \{i \in \{1,\dots,m\} \mid w_{1,i} > 0, w_{2,i} = 0\},$$
$$\mathcal{I}_{0+}(w) = \{i \in \{1,\dots,m\} \mid w_{1,i} = 0, w_{2,i} > 0\},$$
$$\mathcal{I}_{00}(w) = \{i \in \{1,\dots,m\} \mid w_{1,i} = 0, w_{2,i} = 0\}.$$

Tight NLP (TNLP) at w^*

$$\begin{aligned}
\min_{w \in \mathbb{R}^n} \quad & f(w) \\
\text{s.t.} \quad & g(w) = 0, \\
& h(w) \geq 0, \\
& w_{1,i} = 0,\ w_{2,i} \geq 0,\ i \in \mathcal{I}_{0+}(w^*), \\
& w_{1,i} \geq 0,\ w_{2,i} = 0,\ i \in \mathcal{I}_{+0}(w^*), \\
& w_{1,i} = 0,\ w_{2,i} = 0,\ i \in \mathcal{I}_{00}(w^*).
\end{aligned}$$



Pieces of the MPEC: the Tight Nonlinear Program (TNLP)

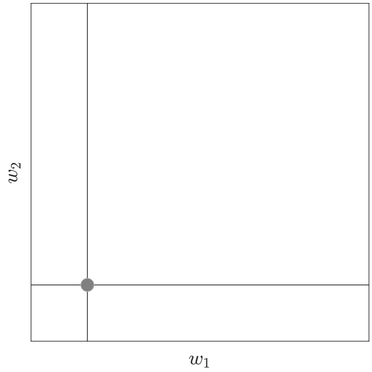
Regular NLPs, used to define MPEC-specific concepts.

MPEC active sets

$$\mathcal{I}_{+0}(w) = \{i \in \{1,\dots,m\} \mid w_{1,i} > 0, w_{2,i} = 0\},$$
$$\mathcal{I}_{0+}(w) = \{i \in \{1,\dots,m\} \mid w_{1,i} = 0, w_{2,i} > 0\},$$
$$\mathcal{I}_{00}(w) = \{i \in \{1,\dots,m\} \mid w_{1,i} = 0, w_{2,i} = 0\}.$$

Tight NLP (TNLP) at w^*

$$\begin{aligned}
\min_{w \in \mathbb{R}^n} \quad & f(w) \\
\text{s.t.} \quad & g(w) = 0, \\
& h(w) \geq 0, \\
& w_{1,i} = 0,\ w_{2,i} \geq 0,\ i \in \mathcal{I}_{0+}(w^*), \\
& w_{1,i} \geq 0,\ w_{2,i} = 0,\ i \in \mathcal{I}_{+0}(w^*), \\
& w_{1,i} = 0,\ w_{2,i} = 0,\ i \in \mathcal{I}_{00}(w^*).
\end{aligned}$$



Pieces of the MPEC: the Relaxed Nonlinear Program (RNLP)

Regular NLPs, used to define MPEC-specific concepts.

MPEC active sets

$$\mathcal{I}_{+0}(w) = \{i \in \{1,\dots,m\} \mid w_{1,i} > 0, w_{2,i} = 0\},$$
$$\mathcal{I}_{0+}(w) = \{i \in \{1,\dots,m\} \mid w_{1,i} = 0, w_{2,i} > 0\},$$
$$\mathcal{I}_{00}(w) = \{i \in \{1,\dots,m\} \mid w_{1,i} = 0, w_{2,i} = 0\}.$$

Pieces of the MPEC: the Relaxed Nonlinear Program (RNLP)

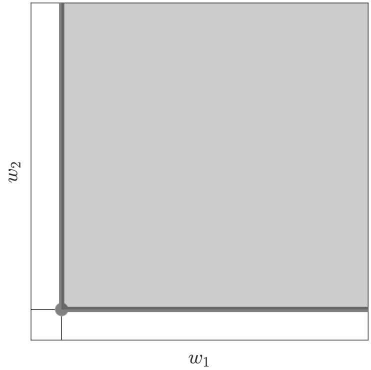
Regular NLPs, used to define MPEC-specific concepts.

MPEC active sets

$$\mathcal{I}_{+0}(w) = \{i \in \{1,\dots,m\} \mid w_{1,i} > 0, w_{2,i} = 0\},$$
$$\mathcal{I}_{0+}(w) = \{i \in \{1,\dots,m\} \mid w_{1,i} = 0, w_{2,i} > 0\},$$
$$\mathcal{I}_{00}(w) = \{i \in \{1,\dots,m\} \mid w_{1,i} = 0, w_{2,i} = 0\}.$$

Relaxed NLP (RNLP) at w^*

$$\begin{aligned}
\min_{w\in\mathbb{R}^n} \quad & f(w) \\
\text{s.t.} \quad & g(w) = 0, \\
& h(w) \geq 0, \\
& w_{1,i} = 0,\ w_{2,i} \geq 0,\ i \in \mathcal{I}_{0+}(w^*), \\
& w_{1,i} \geq 0,\ w_{2,i} = 0,\ i \in \mathcal{I}_{+0}(w^*), \\
& w_{1,i} \geq 0,\ w_{2,i} \geq 0,\ i \in \mathcal{I}_{00}(w^*).
\end{aligned}$$



Pieces of the MPEC: the Branch Nonlinear Program (BNLP)

Regular NLPs, used to define MPEC-specific concepts.

MPEC active sets

$$\mathcal{I}_{+0}(w) = \{i \in \{1,\dots,m\} \mid w_{1,i} > 0, w_{2,i} = 0\},$$
$$\mathcal{I}_{0+}(w) = \{i \in \{1,\dots,m\} \mid w_{1,i} = 0, w_{2,i} > 0\},$$
$$\mathcal{I}_{00}(w) = \{i \in \{1,\dots,m\} \mid w_{1,i} = 0, w_{2,i} = 0\}.$$

Partitioning of the degenerate set $\mathcal{I}_{00}(w)$

$\mathcal{D}_1(w) \cup \mathcal{D}_2(w) = \mathcal{I}_{00}(w)$,
$\mathcal{D}_1(w) \cap \mathcal{D}_2(w) = \emptyset$,
$\mathcal{I}_1(w) := \mathcal{I}_{0+}(w) \cup \mathcal{D}_1(w)$,
$\mathcal{I}_2(w) := \mathcal{I}_{+0}(w) \cup \mathcal{D}_2(w)$,

Pieces of the MPEC: the Branch Nonlinear Program (BNLP)

Regular NLPs, used to define MPEC-specific concepts.

MPEC active sets

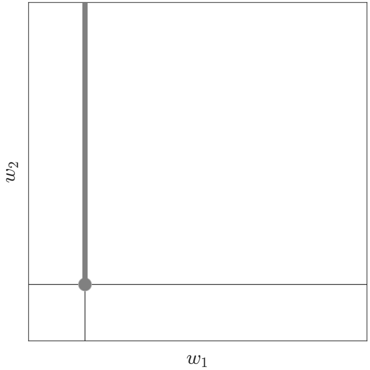
$$\mathcal{I}_{+0}(w) = \{i \in \{1,\ldots,m\} \mid w_{1,i} > 0, w_{2,i} = 0\},$$
$$\mathcal{I}_{0+}(w) = \{i \in \{1,\ldots,m\} \mid w_{1,i} = 0, w_{2,i} > 0\},$$
$$\mathcal{I}_{00}(w) = \{i \in \{1,\ldots,m\} \mid w_{1,i} = 0, w_{2,i} = 0\}.$$

Partitioning of the degenerate set $\mathcal{I}_{00}(w)$

$$\mathcal{D}_1(w) \cup \mathcal{D}_2(w) = \mathcal{I}_{00}(w),$$
$$\mathcal{D}_1(w) \cap \mathcal{D}_2(w) = \emptyset,$$
$$\mathcal{I}_1(w) := \mathcal{I}_{0+}(w) \cup \mathcal{D}_1(w),$$
$$\mathcal{I}_2(w) := \mathcal{I}_{+0}(w) \cup \mathcal{D}_2(w),$$

Branch NLP (BNLP) at w^*

$$\begin{aligned}
\min_{w \in \mathbb{R}^n} \quad & f(w) \\
\text{s.t.} \quad & g(w) = 0, \\
& h(w) \geq 0, \\
& w_{1,i} = 0,\ w_{2,i} \geq 0,\ i \in \mathcal{I}_1(w^*) \\
& w_{1,i} \geq 0,\ w_{2,i} = 0,\ i \in \mathcal{I}_2(w^*).
\end{aligned}$$



Pieces of the MPEC: the Branch Nonlinear Program (BNLP)

Regular NLPs, used to define MPEC-specific concepts.

MPEC active sets

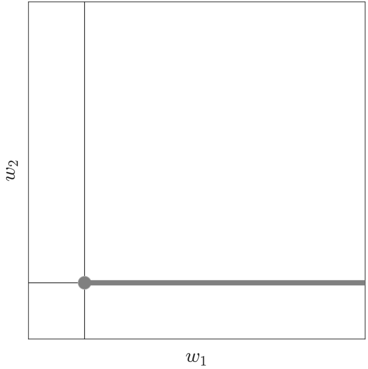
$$\mathcal{I}_{+0}(w) = \{i \in \{1,\dots,m\} \mid w_{1,i} > 0, w_{2,i} = 0\},$$
$$\mathcal{I}_{0+}(w) = \{i \in \{1,\dots,m\} \mid w_{1,i} = 0, w_{2,i} > 0\},$$
$$\mathcal{I}_{00}(w) = \{i \in \{1,\dots,m\} \mid w_{1,i} = 0, w_{2,i} = 0\}.$$

Branch NLP (BNLP) at w^*

$$\begin{aligned} \min_{w\in\mathbb{R}^n} \quad & f(w) \\ \text{s.t.} \quad & g(w) = 0, \\ & h(w) \geq 0, \\ & w_{1,i} = 0,\ w_{2,i} \geq 0,\ i \in \mathcal{I}_1(w^*) \\ & w_{1,i} \geq 0,\ w_{2,i} = 0,\ i \in \mathcal{I}_2(w^*). \end{aligned}$$

Partitioning of the degenerate set $\mathcal{I}_{00}(w)$

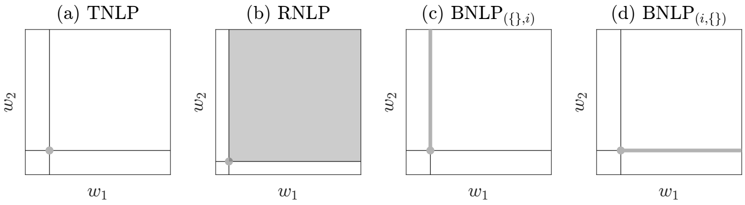
$\mathcal{D}_1(w) \cup \mathcal{D}_2(w) = \mathcal{I}_{00}(w)$,
$\mathcal{D}_1(w) \cap \mathcal{D}_2(w) = \emptyset$,
$\mathcal{I}_1(w) := \mathcal{I}_{0+}(w) \cup \mathcal{D}_1(w)$,
$\mathcal{I}_2(w) := \mathcal{I}_{+0}(w) \cup \mathcal{D}_2(w)$,



Relation between MPEC pieces

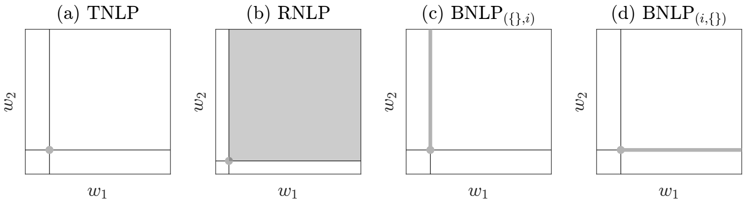
MPEC feasible set relations

$$\Omega_{\mathrm{TNLP}} = \bigcap_{(\mathcal{I}_1,\mathcal{I}_2)} \Omega_{\mathrm{BNLP}(\mathcal{I}_1,\mathcal{I}_2)} \subset \Omega = \bigcup_{(\mathcal{I}_1,\mathcal{I}_2)} \Omega_{\mathrm{BNLP}(\mathcal{I}_1,\mathcal{I}_2)}.$$



MPEC feasible set relations

$$\Omega_{\mathrm{TNLP}} = \bigcap_{(\mathcal{I}_1,\mathcal{I}_2)} \Omega_{\mathrm{BNLP}(\mathcal{I}_1,\mathcal{I}_2)} \subset \Omega = \bigcup_{(\mathcal{I}_1,\mathcal{I}_2)} \Omega_{\mathrm{BNLP}(\mathcal{I}_1,\mathcal{I}_2)}.$$



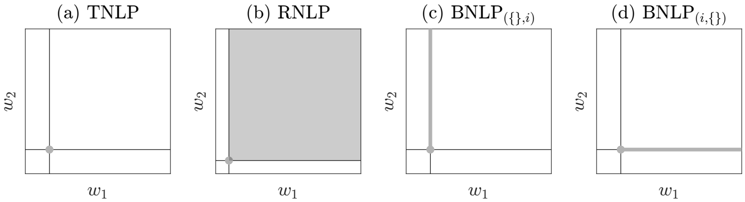
On the solutions:

- If w^* is a local minimizer of the RNLP, then it is a local minimizer of the MPEC. The converse is not true.
- If w^* is a local minimizer of the MPEC then it is a local minimizer of the TNLP.
- The point w^* is a local minimizer of the MPEC if and only if it is a local minimizer of every BNLP.

Relation between MPEC pieces

MPEC feasible set relations

$$\Omega_{\mathrm{TNLP}} = \bigcap_{(\mathcal{I}_1,\mathcal{I}_2)} \Omega_{\mathrm{BNLP}(\mathcal{I}_1,\mathcal{I}_2)} \subset \Omega = \bigcup_{(\mathcal{I}_1,\mathcal{I}_2)} \Omega_{\mathrm{BNLP}(\mathcal{I}_1,\mathcal{I}_2)}.$$



Summary:

- The TNLP, RNLP, and BNLPs are regular nonlinear optimization problems.
- If we know the right TNLP/BNLP, we can just solve a regular NLP to solve the MPEC.
- There are $2^{|\mathcal{I}_{00}|}$ BNLPs - highlighting the combinatorial nature.
- They are used to define MPEC specific definitions, e.g., we say MPEC-LICQ holds at w if standard LICQ holds for the TNLP at w.

Dual stationarity conditions for MPECs

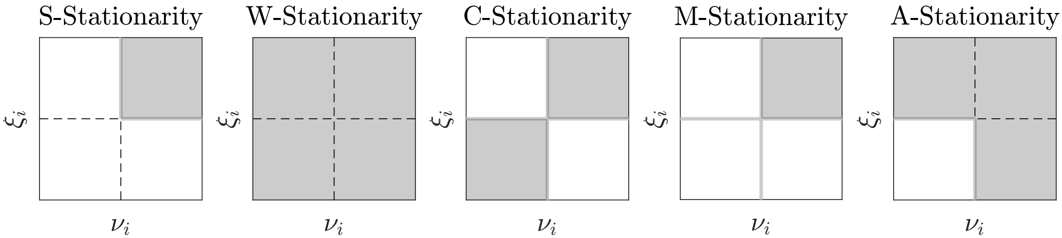
Definition (Stationarity conditions for MPECs)

- Weak Stationarity (W-stationarity): A point $w^* \in \Omega$ is called W-stationary if the corresponding TNLP admits the satisfaction of the KKT conditions, i.e., there exist Lagrange multipliers λ^*, μ^*, ν^* and ξ^* such that:

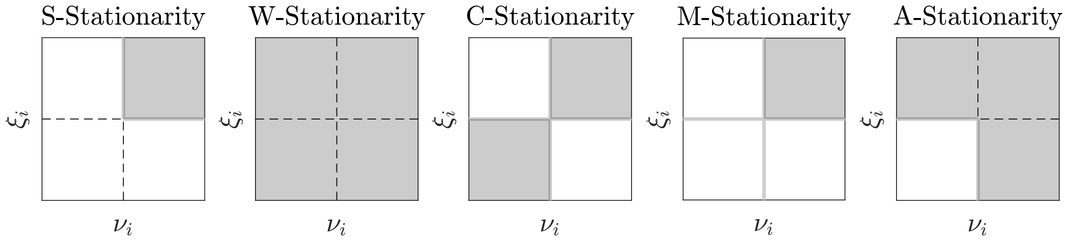
$$
\begin{aligned}
&\nabla_w f(w^*) - \nabla_w g(w^*)\lambda^* - \nabla_w h(w^*)\mu^* - (\nabla_w w_1)\nu^* - (\nabla_w w_2)\xi^* = 0,\\
&g(w^*) = 0,\\
&0 \leq \mu^* \perp h(w^*) \geq 0,\\
&w_{1,i}^* \geq 0, \nu_i^* = 0, \quad \text{for all } i \in \mathcal{I}_{+0}(w^*),\\
&w_{2,i}^* \geq 0, \xi_i^* = 0, \quad \text{for all } i \in \mathcal{I}_{0+}(w^*),\\
&w_{1,i}^* = 0,\ \nu_i^* \in \mathbb{R}, \quad \text{for all } i \in \mathcal{I}_{0+}(w^*) \cup \mathcal{I}_{00}(w^*),\\
&w_{2,i}^* = 0,\ \xi_i^* \in \mathbb{R}, \quad \text{for all } i \in \mathcal{I}_{+0}(w^*) \cup \mathcal{I}_{00}(w^*).
\end{aligned}
$$

- Strong Stationarity (S-stationarity): A point $w^* \in \Omega$ is called S-stationary if it is weakly stationary and $\nu_i^* \geq 0, \xi_i^* \geq 0$ for all $i \in \mathcal{I}_{00}(w^*)$.

Relation between primal and dual stationarity for MPECs



Relation between primal and dual stationarity for MPECs



Theorem (Theorem 4 in [Scheel and Scholtes, 2000])

If w^* is a S-stationary point of the MPEC (5), then it is also B-stationarity. If in addition the MPEC-LICQ holds at w^*, then the reverse implication is also true.

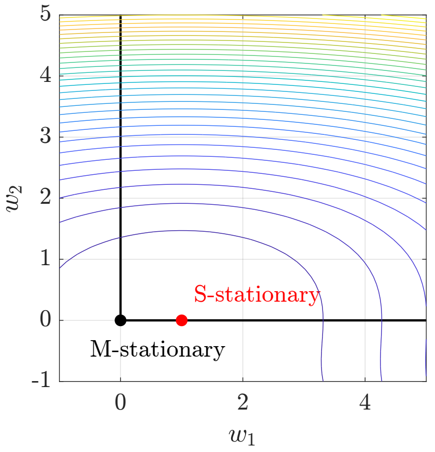
- MPEC-LICQ cannot be relaxed to MPEC-MFCQ. Next weaker concept, M stationary points can be B, but they do not have to.
- Our example from the begging satisfied MPEC-MFCQ and was M-stationary.

M-stationarity allows first-order descent directions

Consider the two-dimensional MPEC:

$$\begin{aligned} \min_{w \in \mathbb{R}^2} \quad & (w_1 - 1)^2 + w_2^2 + w_2^3 \\ \text{s.t.} \quad & 0 \leq w_1 \perp w_2 \geq 0. \end{aligned}$$

- The origin $w^* = 0$ is an M-stationary point with the optimal multipliers $\nu = -2$, $\xi = 0$.

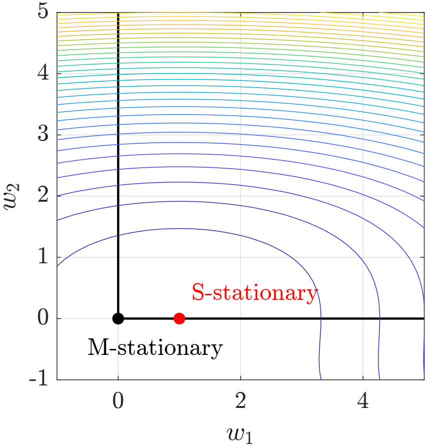


M-stationarity allows first-order descent directions

Consider the two-dimensional MPEC:

$$\begin{aligned}\min_{w\in\mathbb{R}^2} \quad & (w_1-1)^2 + w_2^2 + w_2^3 \\ \text{s.t.} \quad & 0 \le w_1 \perp w_2 \ge 0.\end{aligned}$$

- The origin $w^* = 0$ is an M-stationary point with the optimal multipliers $\nu = -2$, $\xi = 0$.
- There exists a descent direction $d = (1,0)$ with $\nabla f(w^*)^\top d = -2 < 0$.
- The origin is not B-stationary.

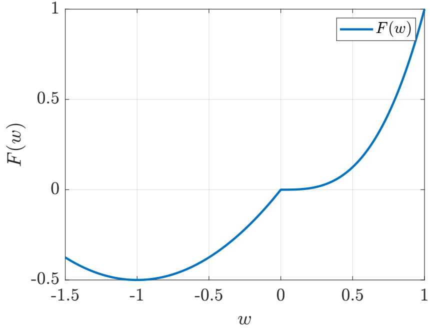


M-stationarity allows first-order descent directions

Consider the two-dimensional MPEC:

$$\min_{w\in\mathbb{R}^2} \quad (w_1-1)^2 + w_2^2 + w_2^3$$
$$\text{s.t.} \quad 0 \le w_1 \perp w_2 \ge 0.$$

- The origin $w^* = 0$ is an M-stationary point with the optimal multipliers $\nu = -2$, $\xi = 0$.
- There exists a descent direction $d = (1, 0)$ with $\nabla f(w^*)^\top d = -2 < 0$.
- The origin is not B-stationary.



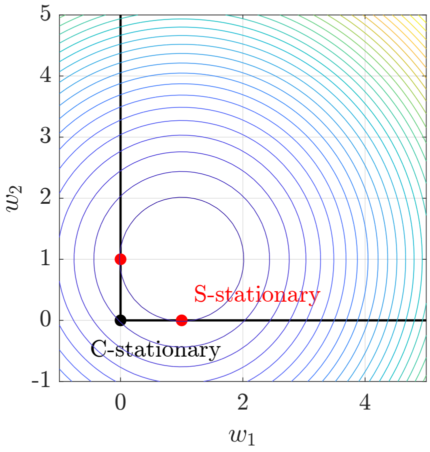
The kink in the example from Lecture 4 corresponds to an M-stationary point.

C-stationarity allows first-order descent directions

Consider the two-dimensional MPEC:

$$\begin{aligned}\min_{w\in\mathbb{R}^2} \quad & (w_1-1)^2+(w_2-1)^2 \\ \text{s.t.} \quad & 0 \leq w_1 \perp w_2 \geq 0.\end{aligned}$$

- The origin $w^* = 0$ is an C-stationary point with the optimal multipliers $\nu = -2$, $\xi = 2$.

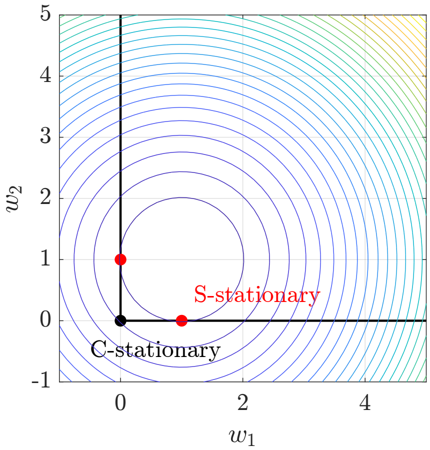


C-stationarity allows first-order descent directions

Consider the two-dimensional MPEC:

$$\begin{aligned} \min_{w \in \mathbb{R}^2} \quad & (w_1 - 1)^2 + (w_2 - 1)^2 \\ \text{s.t.} \quad & 0 \leq w_1 \perp w_2 \geq 0. \end{aligned}$$

- The origin $w^* = 0$ is an C-stationary point with the optimal multipliers $\nu = -2$, $\xi = 2$.
- There exists two descent direction $d = (1, 0)$ and $d = (0, 1)$
- The origin is not B-stationary, $w = (1, 0)$ and $w = (0, 1)$ are S-stationary.



Summary

- Complementarity constraints are a very powerful modeling tool.
- Formulations with disjoint feasible regions should be avoided.
- MPECs violate standard constraint qualifications at all feasible points.
- KKT conditions are not the right tool to identify solution candidates and build algorithm upon.
- Tailored MPEC theory (and algorithms) exploit the piecewise structure.
- Only S-stationarity can verify B-stationarity, all others may allow descent directions = they are **spurious** stationary concepts.

Cited references I

Achtziger, W. and Kanzow, C. (2008).
Mathematical programs with vanishing constraints: optimality conditions and constraint qualifications.
Mathematical Programming, 114:69–99.

Albrecht, S. and Ulbrich, M. (2017).
Mathematical programs with complementarity constraints in the context of inverse optimal control for locomotion.
Optimization Methods and Software, 32(4):670–698.

ApS, M. (2024).
Mosek modeling cookbook.

Baumrucker, B., Renfro, J. G., and Biegler, L. T. (2008).
Mpec problem formulations and solution strategies with chemical engineering applications.
Computers & Chemical Engineering, 32(12):2903–2913.

Cited references II

Baumrucker, B. T. and Biegler, L. T. (2009).
Mpec strategies for optimization of a class of hybrid dynamic systems.
Journal of Process Control, 19(8):1248–1256.

Biegler, L. T. (2010).
Nonlinear Programming.
MOS-SIAM Series on Optimization. SIAM.

Feng, M., Mitchell, J. J., Pang, J. S., Shen, X., and Waechter, A. (2018).
Complementarity formulations of ℓ_0-norm optimization.
Pacific Journal of Optimization, 14(2):273–305.

Flegel, M. L. and Kanzow, C. (2005).
Abadie-type constraint qualification for mathematical programs with equilibrium constraints.
Journal of Optimization Theory and Applications, 124(3):595–614.

Cited references III

Fletcher, R. and Leyffer, S. (2004).
Solving mathematical programs with complementarity constraints as nonlinear programs.
Optimization Methods and Software, 19(1):15–40.

Guo, L. and Ye, J. J. (2016).
Necessary optimality conditions for optimal control problems with equilibrium constraints.
SIAM Journal on Control and Optimization, 54(5):2710–2733.

Hegerhorst-Schultchen, L. C., Kirches, C., and Steinbach, M. C. (2020).
On the relation between mpecs and optimization problems in abs-normal form.
Optimization Methods and Software, 35(3):560–575.

Hoheisel, T., Kanzow, C., and Schwartz, A. (2013).
Theoretical and numerical comparison of relaxation methods for mathematical programs with complementarity constraints.
Mathematical Programming, 137(1):257–288.

Cited references IV

Hu, J., Mitchell, J. E., Pang, J.-S., and Yu, B. (2012).
On linear programs with linear complementarity constraints.
Journal of Global Optimization, 53:29–51.

Kanzow, C. and Schwartz, A. (2015).
The price of inexactness: convergence properties of relaxation methods for mathematical programs with complementarity constraints revisited.
Mathematics of Operations Research, 40(2):253–275.

Kanzow, C., Schwartz, A., and Weiß, F. (2024).
The sparse (st) optimization problem: Reformulations, optimality, stationarity, and numerical results.
Computational Optimization and Applications, pages 1–36.

Kim, Y., Leyffer, S., and Munson, T. (2020).
Mpec methods for bilevel optimization problems.
In *Bilevel Optimization*, pages 335–360. Springer.

Cited references V

Krug, R., Mehrmann, V., and Schmidt, M. (2021).
Nonlinear optimization of district heating networks.
Optimization and Engineering, 22(2):783–819.

Luo, Z., Pang, J., and Ralph, D. (1996).
Mathematical Programs with Equilibrium Constraints.
Cambridge University Press, Cambridge.

Nurkanović, A. (2023).
Numerical Methods for Optimal Control of Nonsmooth Dynamical Systems.
PhD thesis, University of Freiburg.

Nurkanović, A., Pozharskiy, A., and Diehl, M. (2024).
Solving mathematical programs with complementarity constraints arising in nonsmooth optimal control.
Vietnam Journal of Mathematics, pages 1–39.

Cited references VI

Pozharskiy, A., Nurkanović, A., and Diehl, M. (2024).
Finite elements with switch detection for numerical optimal control of projected dynamical systems.
arXiv preprint arXiv:2404.05367.

Raghunathan, A. and Biegler, L. (2003).
Mathematical programs with equilibrium constraints (MPECs) in process engineering.
Computers and Chemical Engineering, 27:1381–1392.

Scheel, H. and Scholtes, S. (2000).
Mathematical programs with complementarity constraints: Stationarity, optimality, and sensitivity.
Math. Oper. Res., 25:1–22.

Szmuk, M., Reynolds, T. P., and Açıkmeşe, B. (2020).
Successive convexification for real-time six-degree-of-freedom powered descent guidance with state-triggered constraints.
Journal of Guidance, Control, and Dynamics, 43(8):1399–1413.

Cited references VII

Vieira, A., Brogliato, B., and Prieur, C. (2019).
Quadratic optimal control of linear complementarity systems: First-order necessary conditions and numerical analysis.
IEEE Transactions on Automatic Control, 65(6):2743–2750.

Wächter, A. and Biegler, L. T. (2006).
On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming.
Mathematical Programming, 106(1):25–57.

Wensing, P. M., Posa, M., Hu, Y., Escande, A., Mansard, N., and Del Prete, A. (2023).
Optimization-based control for dynamic legged robots.
IEEE Transactions on Robotics.

State triggered constraints (vanishing equality constraints) - Flipped

Introduced in [Szmuk et al., 2020]

State triggered constraints

If $H_i(w) > 0$ then $G_i(w) = 0$, otherwise if $H_i(w) \leq 0$ then $G_i(w) \in \mathbb{R}$

State triggered constraints (vanishing equality constraints) - Flipped

Introduced in [Szmuk et al., 2020]

State triggered constraints

If $H_i(w) > 0$ then $G_i(w) = 0$, otherwise if $H_i(w) \leq 0$ then $G_i(w) \in \mathbb{R}$

- Logical formulation:

$$H_i(w) > 0 \implies G_i(w) = 0, \quad i = 1, \ldots, m.$$

State triggered constraints (vanishing equality constraints) - Flipped

Introduced in [Szmuk et al., 2020]

State triggered constraints

If $H_i(w) > 0$ then $G_i(w) = 0$, otherwise if $H_i(w) \leq 0$ then $G_i(w) \in \mathbb{R}$

- Logical formulation:

$$H_i(w) > 0 \implies G_i(w) = 0, \quad i = 1, \ldots, m.$$

- Complementarity formulation:

$$\begin{aligned} z &\perp G(w), \\ 0 \leq -H(w) + z &\perp z \geq 0. \end{aligned}$$

State triggered constraints

If $H_i(w) > 0$ then $G_i(w) = 0$, otherwise if $H_i(w) \leq 0$ then $G_i(w) \in \mathbb{R}$

- Logical formulation:

$$H_i(w) > 0 \implies G_i(w) = 0, \quad i = 1, \ldots, m.$$

- Complementarity formulation:

$$\begin{aligned} z \perp G(w), \\ 0 \leq -H(w) + z \perp z \geq 0. \end{aligned}$$

- Interpretation:
  - $H_i(w) > 0 \implies z_i = H_i(w) > 0 \implies G_i(w) = 0,$
  - $H_i(w) < 0 \implies z_i = 0 \implies G_i(w) \in \mathbb{R},$
  - $H_i(w) = 0 \implies z_i = 0 \implies G_i(w) \in \mathbb{R}.$

State triggered inequality constraints

Introduced in [Szmuk et al., 2020]

State triggered inequality constraints

If $H_i(w) > 0$ then $G_i(w) \geq 0$, otherwise if $H_i(w) \leq 0$ then $G_i(w) \in \mathbb{R}$

State triggered inequality constraints

Introduced in [Szmuk et al., 2020]

State triggered inequality constraints

If $H_i(w) > 0$ then $G_i(w) \geq 0$, otherwise if $H_i(w) \leq 0$ then $G_i(w) \in \mathbb{R}$

- Logical formulation:

$$H_i(w) > 0 \implies G_i(w) = 0, \quad i = 1, \ldots, m.$$

State triggered inequality constraints

Introduced in [Szmuk et al., 2020]

State triggered inequality constraints

If $H_i(w) > 0$ then $G_i(w) \geq 0$, otherwise if $H_i(w) \leq 0$ then $G_i(w) \in \mathbb{R}$

- Logical formulation:

$$H_i(w) > 0 \implies G_i(w) = 0, \quad i = 1, \dots, m.$$

- Complementarity formulation:

$$\begin{aligned} zG(w) &\geq 0, \\ 0 \leq -H(w) + z \perp z &\geq 0. \end{aligned}$$

State triggered inequality constraints

Introduced in [Szmuk et al., 2020]

State triggered inequality constraints

If $H_i(w) > 0$ then $G_i(w) \geq 0$, otherwise if $H_i(w) \leq 0$ then $G_i(w) \in \mathbb{R}$

- Logical formulation:

$$H_i(w) > 0 \implies G_i(w) = 0, \quad i = 1, \ldots, m.$$

- Complementarity formulation:

$$\begin{aligned} zG(w) &\geq 0, \\ 0 \leq -H(w) + z \perp z &\geq 0. \end{aligned}$$

- Interpretation:
  - $H_i(w) > 0 \implies z_i = H_i(w) > 0 \implies G_i(w) \geq 0$,
  - $H_i(w) < 0 \implies z_i = 0 \implies G_i(w) \in \mathbb{R}$,
  - $H_i(w) = 0 \implies z_i = 0 \implies G_i(w) \in \mathbb{R}$.