Technische Universität Braunschweig

Mathematical Programs with Complementarity Constraints Part 1: Theory

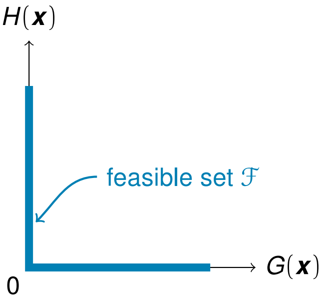
Christian Kirches | Freiburg | September 14, 2023

Overview

- Problem Classes
- KKT Theorem and Details
- Constraint Qualifications
- Optimality Conditions

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.

Smooth-looking Multiplicative Formulations

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

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

The problem may also be stated with a non-smooth constraint:

- $\min\{\boldsymbol{u}, \boldsymbol{v}\} = \mathbf{0}$
- $\min\{u_i, v_i\} = 0$ for all $1 \leqslant i \leqslant c$

Vertical Form

Any MPCC can be cast in the so-called **vertical form**, using only **orthogonal** complementarities:

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

When solving MPCCs numerically (later), the vertical form guarantees linear feasibility and typically shows better convergence behavior.

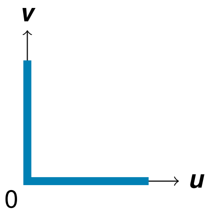
Lifted Form

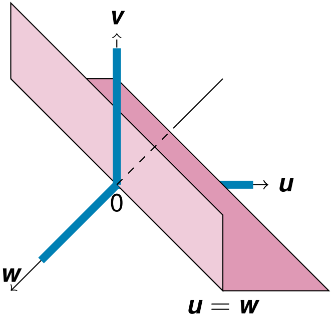
Any MPCC can be cast in a **lifted form** by introducing

- a slack $\boldsymbol{w} \in \mathbb{R}^c$,
- a penalty function $p(\boldsymbol{w})$,
- and a penalty parameter $\pi > 0$:

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

Example: $p(\boldsymbol{w}) = \|\boldsymbol{w}\|_1$





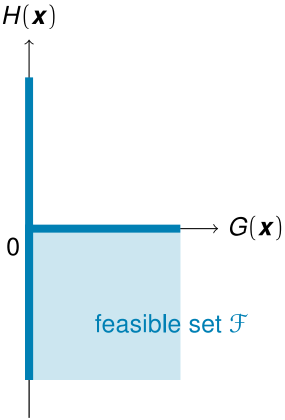
Math. Programs with Vanishing Constraints

MPCCs have a close relative, MPVCs:

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

Any MPVC can be cast as an MPCC by introduction of a slack vector $\boldsymbol{s}$:

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



In a solution, the slack will be degenerate. A more detailed analysis shows that MPVCs are *slightly* more regular than an MPCC plus a slack vector.

Equilibrium Constraints (MPECs)

$$\boxed{\begin{aligned} &\min_{\boldsymbol{x}} \; F(\boldsymbol{x}, \boldsymbol{y}) \\ &\text{ s.t. } \boldsymbol{x} \in \mathcal{X}, \;\; \boldsymbol{y} \in S(\boldsymbol{x}) \end{aligned}}$$

wherein $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is the objective and $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is a set-valued map, called the “equilibrium constraint”.

One example: **Bi-level programs** The set

$$S(\boldsymbol{x}) = \underset{\boldsymbol{y}}{\operatorname{argmin}} \left\{ F(\boldsymbol{x}, \boldsymbol{y}) \;\middle|\; \boldsymbol{y} \in \mathcal{Y} \right\}$$

is, for a given vector $\boldsymbol{x}$, the solution set of the inner problem

$$\boxed{\min_{\boldsymbol{y}} \;\; F(\boldsymbol{x}, \boldsymbol{y}) \text{ s.t. } \boldsymbol{y} \in \mathcal{Y}(\boldsymbol{x})}$$

Under assumptions, the inner problem may be replaced by its first order necessary conditions. We obtain an MPCC if $\mathcal{Y}(\boldsymbol{x})$ contains inequality constraints.

Game Theory

Stackelberg game: Asymmetric two-player game over turns $k \geqslant 1$.

Leader controls $\boldsymbol{x}$ and minimizes $L(\boldsymbol{x}, \boldsymbol{y})$ considering set $S(\boldsymbol{x})$ of follower's responses:

(leader)

$$\boxed{\begin{aligned} &\min_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{y}) \\ &\text{ s.t. } \boldsymbol{x} \in \mathcal{X},\ \boldsymbol{y} \in S(\boldsymbol{x}) \end{aligned}}$$

$$\Updownarrow$$

(follower)

$$\boxed{\begin{aligned} &\min_{\boldsymbol{y}^{(k)}} F(\boldsymbol{x}, \boldsymbol{y}^{(k)}, \boldsymbol{y}^{(k-1)}) \\ &\text{ s.t. } \boldsymbol{y}^{(k)} \in \mathcal{Y}(\boldsymbol{x}, \boldsymbol{y}^{(k-1)}) \end{aligned}}$$

Follower controls $\boldsymbol{y}^{(k)}$ given the leader's choice $\boldsymbol{x}$ and the follower's response $\boldsymbol{y}^{(k-1)}$ in the previous turn.

Game Theory

For a given element $\boldsymbol{x}$ assume

$$\mathcal{Y}(\boldsymbol{x}, \boldsymbol{y}^{(k-1)}) = \left\{ \boldsymbol{y} \;\middle|\; G(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{0},\ H(\boldsymbol{x}, \boldsymbol{y}) \geqslant \boldsymbol{0} \right\}.$$

Under a suitable constraint qualification, an element

$$\boldsymbol{y} \in S(\boldsymbol{x}) = \operatorname*{argmin}_{\boldsymbol{y}} \left\{ F(\boldsymbol{x}, \boldsymbol{y}) \;\middle|\; \boldsymbol{y} \in \mathcal{Y}(\boldsymbol{x}, \boldsymbol{y}^{(k-1)}) \right\}$$

necessarily satisfies

$$\begin{aligned}
\nabla_{\boldsymbol{y}} F(\boldsymbol{x}, \boldsymbol{y}) + \nabla_{\boldsymbol{y}} G(\boldsymbol{x}, \boldsymbol{y}) \cdot \boldsymbol{\lambda} + \nabla_{\boldsymbol{y}} H(\boldsymbol{x}, \boldsymbol{y}) \cdot \boldsymbol{\mu} &= \boldsymbol{0} \\
G(\boldsymbol{x}, \boldsymbol{y}) &= \boldsymbol{0} \\
\boldsymbol{0} \leqslant \boldsymbol{\mu} \perp H(\boldsymbol{x}, \boldsymbol{y}) &\geqslant \boldsymbol{0}
\end{aligned}$$

for some vectors $\boldsymbol{\lambda}$, $\boldsymbol{\mu}$. Under assumptions, we may replace the constraint $\boldsymbol{y} \in S(\boldsymbol{x})$ in the leader's problem by these necessary conditions.

Nonconvex Relaxations of Discrete Problems

MINLP with indicator constraints $G_i(\boldsymbol{x}) \geqslant \mathbf{0}$ on indicators variables ω_i:

$$\boxed{\begin{aligned} \min_{\boldsymbol{x},\boldsymbol{\omega}} \; & F(\boldsymbol{x},\boldsymbol{\omega}) \\ \text{s.t. } & C(\boldsymbol{x}) = \mathbf{0}, \;\; D(\boldsymbol{x}) \geqslant \mathbf{0} \\ & \omega_i \cdot G_i(\boldsymbol{x}) \geqslant \mathbf{0} \\ & \mathbf{1}^T\boldsymbol{\omega} = 1, \; \omega_i \in \{0,1\}, \;\; 1 \leqslant i \leqslant n_\omega \end{aligned}}$$

The problem admits a non-convex relaxation, which is an MPVC:

$$\boxed{\begin{aligned} \min_{\boldsymbol{x},\boldsymbol{\alpha}} \; & F(\boldsymbol{x},\boldsymbol{\alpha}) \\ \text{s.t. } & C(\boldsymbol{x}) = \mathbf{0}, \;\; D(\boldsymbol{x}) \geqslant \mathbf{0} \\ & \alpha_i \cdot G_i(\boldsymbol{x}) \geqslant \mathbf{0} \\ & \mathbf{1}^T\boldsymbol{\alpha} = 1, \;\; \alpha_i \in [0,1], \;\; 1 \leqslant i \leqslant n_\omega \end{aligned}}$$

Not a magic bullet to combinatorial optimization. Stationary points of the relaxation sometimes yield good initial guesses.

Abs-Normal Form: Structured Nonsmoothness

A non-smooth function $\phi(\boldsymbol{x})$ is in **abs-normal form** if

$$\boxed{\begin{aligned} \phi(\boldsymbol{x}) &= f(\boldsymbol{x}, |\boldsymbol{z}|) \\ \boldsymbol{z} &= F(\boldsymbol{x}, |\boldsymbol{z}|) \quad \partial_{|\boldsymbol{z}|} F \text{ strictly lower triangular} \end{aligned}}$$

such that $z_1 = F(\boldsymbol{x})$ and $z_k = F(\boldsymbol{x}, |z_1|, \ldots, |z_{k-1}|)$ for $k > 1$.

Abs-Normal forms are amenable to automatic differentiation, e.g. `ADOL-C`.

Abs-normal forms are identical to their counterpart complementarity problems in vertical form:

$$\boxed{\begin{aligned} \phi(\boldsymbol{x}) &= f(\boldsymbol{x}, \boldsymbol{u} + \boldsymbol{v}) \\ \boldsymbol{u} - \boldsymbol{v} &= F(\boldsymbol{x}, \boldsymbol{u} + \boldsymbol{v}) \\ \boldsymbol{0} \leqslant \boldsymbol{u} \ &\perp \ \boldsymbol{v} \geqslant \boldsymbol{0} \end{aligned}}$$

Why MPCCs mean Trouble

Example:

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^3}\ & x_1 + x_2 - x_3 \\ \text{s.t.}\ & -4x_1 + x_3 \leqslant 0 && \mid \mu_1 \\ & -4x_2 + x_3 \leqslant 0 && \mid \mu_2 \\ & 0 \leqslant x_1,\ 0 \leqslant x_2,\ x_1 \cdot x_2 = 0 && \mid \mu_3, \mu_4, \lambda \end{aligned}$$

Global Minimum: Observe $x_3 \leqslant 4\min\{x_1, x_2\} = 0$, hence $\boldsymbol{x}^* = (0, 0, 0)$.

Why MPCCs mean Trouble

Example:

$$\begin{array}{lll} \min\limits_{\boldsymbol{x}\in\mathbb{R}^3} & x_1 + x_2 - x_3 & \\ \text{s.t.} & -4x_1 + x_3 \leqslant 0 & \mid \mu_1 \\ & -4x_2 + x_3 \leqslant 0 & \mid \mu_2 \\ & 0 \leqslant x_1,\ 0 \leqslant x_2,\ x_1 \cdot x_2 = 0 & \mid \mu_3, \mu_4, \lambda \end{array}$$

Global Minimum: Observe $x_3 \leqslant 4\min\{x_1, x_2\} = 0$, hence $\boldsymbol{x}^* = (0, 0, 0)$.

Remember the KKT theorem and try to verify stationarity:

$$\begin{pmatrix} -4 & \cdot & -1 & \cdot & \cdot \\ \cdot & -4 & \cdot & -1 & \cdot \\ 1 & 1 & \cdot & \cdot & \cdot \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \\ \lambda \end{pmatrix} = - \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$$

Result: $\mu_3 = 1 - 4\mu_1 \geqslant 0$, $\mu_4 = 1 - 4\mu_2 \geqslant 0$,
$\mu_1 + \mu_2 = 1$, $\mu_1 \geqslant 0$, $\mu_2 \geqslant 0$.

This is impossible, so the global minimizer x^* apparently is not a KKT point!

Why MPCCs are Trouble

Example:

$$\boxed{\min_{\boldsymbol{x}\in\mathbb{R}^2} \psi(\boldsymbol{x}) \text{ s.t. } 0 \leqslant x_1,\ 0 \leqslant x_2,\ x_1 \cdot x_2 = 0}$$

Observation:

- If $x_1 > 0$, $x_2 = 0$ then:
  Gradients of active constraints $(0, 1)^T$ and $(0, x_1)^T$ linearly dependent
- If $x_1 = 0$, $x_2 > 0$ then:
  Gradients of active constraints $(1, 0)^T$ and $(x_2, 0)^T$ linearly dependent
- If $x_1 = 0$, $x_2 = 0$ then:
  Gradients of active constraints $(0, 1)^T$, $(0, 1)^T$, $(0, 0)^T$ linearly dependent

$\Longrightarrow$ Lack of Linear Independence Constraint Qualification!

Indeed: Mangasarian-Fromovitz CQ and Abadie's CQ also don't hold.
Hence, the KKT theorem does not hold.

Basics on Cones

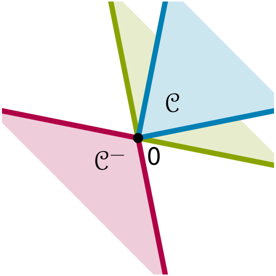
Cones:

- $\mathcal{C} \subseteq \mathbb{R}^n$ **a cone** if $\alpha \boldsymbol{x} \in \mathcal{C}$ for all $\boldsymbol{x} \in \mathcal{C}$ and all real $\alpha \geqslant 0$

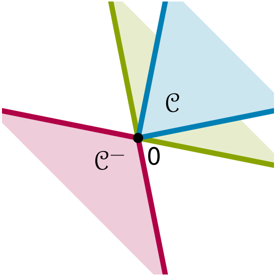
Given cone $\mathcal{C} \subseteq \mathbb{R}^n$,

- $\mathcal{C}^+ := \{\boldsymbol{d} \mid \boldsymbol{d}^T \boldsymbol{x} \geqslant 0 \; \forall \boldsymbol{x} \in \mathcal{C}\}$ is the **dual cone** and

- $\mathcal{C}^- := \{\boldsymbol{d} \mid \boldsymbol{d}^T \boldsymbol{x} \leqslant 0 \; \forall \boldsymbol{x} \in \mathcal{C}\}$ is the **polar cone**



Basics on Cones

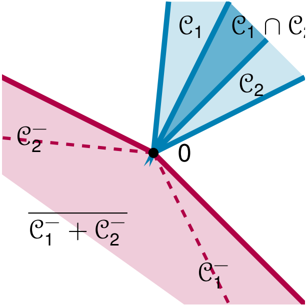


Cones:

- $\mathcal{C} \subseteq \mathbb{R}^n$ **a cone** if $\alpha \boldsymbol{x} \in \mathcal{C}$ for all $\boldsymbol{x} \in \mathcal{C}$ and all real $\alpha \geqslant 0$

Given cone $\mathcal{C} \subseteq \mathbb{R}^n$,

- $\mathcal{C}^+ := \{\boldsymbol{d} \mid \boldsymbol{d}^T \boldsymbol{x} \geqslant 0 \; \forall \boldsymbol{x} \in \mathcal{C}\}$ is the **dual cone** and
- $\mathcal{C}^- := \{\boldsymbol{d} \mid \boldsymbol{d}^T \boldsymbol{x} \leqslant 0 \; \forall \boldsymbol{x} \in \mathcal{C}\}$ is the **polar cone**



Identities:

- $\mathcal{C}^- = \overline{\text{conv}\mathcal{C}}^-$ and $(\mathcal{C}^-)^- = \overline{\text{conv}\mathcal{C}}$
- $(\mathcal{C}_1 \cap \mathcal{C}_2)^- = \overline{\mathcal{C}_1^- + \mathcal{C}_2^-}$

Tangent Cone

Tangent:

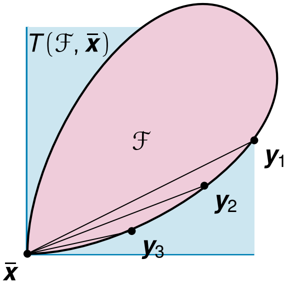
$\boldsymbol{d} \in \mathbb{R}^n$ is **tangent** to $\mathcal{F}$ at $\bar{\boldsymbol{x}}$ if there is a sequence $\{\boldsymbol{y}_k\} \subset \mathcal{F}$ with $\lim_{k\to\infty} \boldsymbol{y}_k \to \bar{\boldsymbol{x}}$ and a sequence $\{t_k\} \subset \mathbb{R}_{\geqslant 0}$ with $\lim_{k\to\infty} t_k = 0$ such that

$$\lim_{k\to\infty} t_k(\boldsymbol{y}_k - \bar{\boldsymbol{x}}) = \boldsymbol{d}.$$

Tangent Cone:

$T(\mathcal{F}, \bar{\boldsymbol{x}}) = \{\boldsymbol{d} \in \mathbb{R}^n \mid \boldsymbol{d}$ tangent to $\mathcal{F}$ at $\bar{\boldsymbol{x}}\}$ is the **tangent cone.**

$P(\mathcal{F}, \bar{\boldsymbol{x}}) = \overline{\operatorname{conv} T(\mathcal{F}, \bar{\boldsymbol{x}})} = T(\mathcal{F}, \bar{\boldsymbol{x}})^{--}$ is the **pseudotangent cone.**



1st Order Necessary Optimality Conditions

Consider the problem

$$\boxed{\min_{\boldsymbol{x} \in \mathbb{R}^n} \ \psi(\boldsymbol{x}) \text{ s.t. } \boldsymbol{x} \in \mathcal{F}}$$

with feasible set $\mathcal{F} \subset \mathbb{R}^n$.

Theorem (1st Order Necessary Optimality Condition)

Let $\bar{\boldsymbol{x}}$ minimize ψ over $\mathcal{F}$. Then

$$\nabla\psi(\bar{\boldsymbol{x}}) \in P^{+}(\mathcal{F}, \bar{\boldsymbol{x}}) := \left\{ \boldsymbol{q} \in \mathbb{R}^n \,\middle|\, \boldsymbol{q}^T \boldsymbol{d} \geqslant 0 \ \forall \boldsymbol{d} \in P(\mathcal{F}, \bar{\boldsymbol{x}}) \right\}$$

holds, and $P(\mathcal{F}, \bar{\boldsymbol{x}}) = \overline{\operatorname{conv} T(\mathcal{F}, \bar{\boldsymbol{x}})}$ denotes the pseudotangent cone of $\mathcal{F}$ at $\bar{\boldsymbol{x}}$.

This theorem is great because we don't have to impose particular structural restrictions on the set $\mathcal{F}$.

Proof (1^{st} Order Necessary Optimality Condition)

Proof:

1. Let $d \in T(\mathcal{F}, \bar{\boldsymbol{x}})$. Then by definition, there is $\{\boldsymbol{x}_k\}_k \subset \mathcal{F}$ with $\lim_{k\to\infty} \boldsymbol{x}_k = \bar{\boldsymbol{x}}$ and $\{t_k\}_k \subset \mathbb{R}_{>0}$ with $\lim_{k\to\infty} t_k = 0$ and $\lim_{k\to\infty} t_k(\boldsymbol{x}_k - \bar{\boldsymbol{x}}) = d$.

Proof (1st Order Necessary Optimality Condition)

Proof:

1. Let $d \in T(\mathcal{F}, \bar{\boldsymbol{x}})$. Then by definition, there is $\{\boldsymbol{x}_k\}_k \subset \mathcal{F}$ with $\lim\limits_{k\to\infty} \boldsymbol{x}_k = \bar{\boldsymbol{x}}$ and $\{t_k\}_k \subset \mathbb{R}_{>0}$ with $\lim\limits_{k\to\infty} t_k = 0$ and $\lim\limits_{k\to\infty} t_k(\boldsymbol{x}_k - \bar{\boldsymbol{x}}) = d$.
2. As $\bar{\boldsymbol{x}}$ minimizes ψ over $\mathcal{F}$, $\psi(\boldsymbol{x}_k) - \psi(\bar{\boldsymbol{x}}) \geqslant 0$ for all $k \geqslant 0$ holds.

Proof (1st Order Necessary Optimality Condition)

Proof:

1. Let $d \in T(\mathcal{F}, \bar{\boldsymbol{x}})$. Then by definition, there is $\{\boldsymbol{x}_k\}_k \subset \mathcal{F}$ with $\lim_{k\to\infty} \boldsymbol{x}_k = \bar{\boldsymbol{x}}$ and $\{t_k\}_k \subset \mathbb{R}_{>0}$ with $\lim_{k\to\infty} t_k = 0$ and $\lim_{k\to\infty} t_k(\boldsymbol{x}_k - \bar{\boldsymbol{x}}) = d$.
2. As $\bar{\boldsymbol{x}}$ minimizes ψ over $\mathcal{F}$, $\psi(\boldsymbol{x}_k) - \psi(\bar{\boldsymbol{x}}) \geqslant 0$ for all $k \geqslant 0$ holds.
3. By differentiability of ψ at $\bar{\boldsymbol{x}}$,

$$\psi(\boldsymbol{x}_k) - \psi(\bar{\boldsymbol{x}}) = \nabla\psi(\bar{\boldsymbol{x}})^T(\boldsymbol{x}_k - \bar{\boldsymbol{x}}) + o(\|\boldsymbol{x}_k - \bar{\boldsymbol{x}}\|).$$

Proof (1^{st} Order Necessary Optimality Condition)

Proof:

1. Let $d \in T(\mathcal{F}, \bar{\boldsymbol{x}})$. Then by definition, there is $\{\boldsymbol{x}_k\}_k \subset \mathcal{F}$ with $\lim\limits_{k\to\infty} \boldsymbol{x}_k = \bar{\boldsymbol{x}}$ and $\{t_k\}_k \subset \mathbb{R}_{>0}$ with $\lim\limits_{k\to\infty} t_k = 0$ and $\lim\limits_{k\to\infty} t_k(\boldsymbol{x}_k - \bar{\boldsymbol{x}}) = d$.
2. As $\bar{\boldsymbol{x}}$ minimizes ψ over $\mathcal{F}$, $\psi(\boldsymbol{x}_k) - \psi(\bar{\boldsymbol{x}}) \geqslant 0$ for all $k \geqslant 0$ holds.
3. By differentiability of ψ at $\bar{\boldsymbol{x}}$,

$$\psi(\boldsymbol{x}_k) - \psi(\bar{\boldsymbol{x}}) = \nabla\psi(\bar{\boldsymbol{x}})^T(\boldsymbol{x}_k - \bar{\boldsymbol{x}}) + o(||\boldsymbol{x}_k - \bar{\boldsymbol{x}}||).$$

4. Then, we have

$$\nabla\psi(\bar{\boldsymbol{x}})^T t_k(\boldsymbol{x}_k - \bar{\boldsymbol{x}}) \geqslant -\frac{o(||\boldsymbol{x}_k - \bar{\boldsymbol{x}}||)}{||\boldsymbol{x}_k - \bar{\boldsymbol{x}}||} t_k||\boldsymbol{x}_k - \bar{\boldsymbol{x}}||.$$

Proof (1^{st} Order Necessary Optimality Condition)

Proof:

1. Let $d \in T(\mathcal{F}, \bar{\boldsymbol{x}})$. Then by definition, there is $\{\boldsymbol{x}_k\}_k \subset \mathcal{F}$ with $\lim\limits_{k\to\infty} \boldsymbol{x}_k = \bar{\boldsymbol{x}}$ and $\{t_k\}_k \subset \mathbb{R}_{>0}$ with $\lim\limits_{k\to\infty} t_k = 0$ and $\lim\limits_{k\to\infty} t_k(\boldsymbol{x}_k - \bar{\boldsymbol{x}}) = d$.
2. As $\bar{\boldsymbol{x}}$ minimizes ψ over $\mathcal{F}$, $\psi(\boldsymbol{x}_k) - \psi(\bar{\boldsymbol{x}}) \geqslant 0$ for all $k \geqslant 0$ holds.
3. By differentiability of ψ at $\bar{\boldsymbol{x}}$,
$$\psi(\boldsymbol{x}_k) - \psi(\bar{\boldsymbol{x}}) = \nabla\psi(\bar{\boldsymbol{x}})^T(\boldsymbol{x}_k - \bar{\boldsymbol{x}}) + o(\|\boldsymbol{x}_k - \bar{\boldsymbol{x}}\|).$$
4. Then, we have
$$\nabla\psi(\bar{\boldsymbol{x}})^T t_k(\boldsymbol{x}_k - \bar{\boldsymbol{x}}) \geqslant -\frac{o(\|\boldsymbol{x}_k - \bar{\boldsymbol{x}}\|)}{\|\boldsymbol{x}_k - \bar{\boldsymbol{x}}\|} t_k\|\boldsymbol{x}_k - \bar{\boldsymbol{x}}\|.$$
5. Now let $k \to \infty$ and obtain
$$\nabla\psi(\bar{\boldsymbol{x}})^T d \geqslant 0. \quad \Box$$

Some more structure

If don't know anything about $\mathcal{F}$, the condition $\boldsymbol{q} \in P^+(\mathcal{F}, \bar{\boldsymbol{x}})$ is difficult to check **computationally**.

Hence, we impose slightly more structure by considering the problem

$$\boxed{\min_{\boldsymbol{x}\in\mathbb{R}^n} \; \psi(\boldsymbol{x}) \text{ s.t. } \boldsymbol{x} \in \mathcal{C}, \; \boldsymbol{a}(\boldsymbol{x}) \in \mathcal{B},}$$

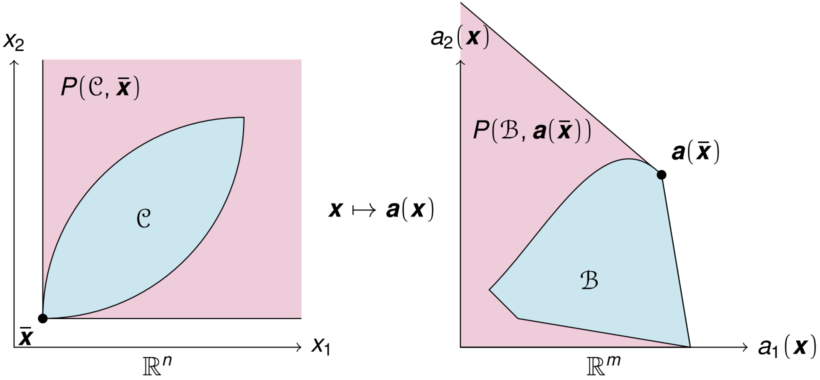
where now $\mathcal{F} = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{x} \in \mathcal{C}, \; \boldsymbol{a}(\boldsymbol{x}) \in \mathcal{B}\}$ and $\boldsymbol{a} : \mathbb{R}^n \to \mathbb{R}^m$, $\mathcal{B} \subset \mathbb{R}^m$.

The sets $\mathcal{B}$, $\mathcal{C}$ are assumed to be **easy** enough to check membership in cones, e.g. by looking at signs of some vector entries.

Cones: $P(\mathcal{F}, \bar{\boldsymbol{x}})$, $P(\mathcal{B}, \boldsymbol{a}(\bar{\boldsymbol{x}}))$

MPCCs: We'll try to encode complementarities in $\mathcal{B}$, $\mathcal{C}$, $\boldsymbol{a}(\boldsymbol{x})$ in a moment.

Some more structure



Guignard's KKT Theorem

Denote by

$$K := \left\{ \boldsymbol{d} \in \mathbb{R}^n \mid \nabla \boldsymbol{a}(\bar{\boldsymbol{x}})^T \boldsymbol{d} \in P(\mathcal{B}, \boldsymbol{a}(\bar{\boldsymbol{x}})) \right\}$$

the cone of first order feasible directions at $\bar{\boldsymbol{x}}$ w.r.t. $\boldsymbol{a}(\bar{\boldsymbol{x}}) \in \mathcal{B}$,
and denote by

$$H := \left\{ \boldsymbol{q} \in \mathbb{R}^n \mid \boldsymbol{q} = \nabla \boldsymbol{a}(\bar{\boldsymbol{x}}) \cdot \boldsymbol{\lambda},\ \boldsymbol{\lambda} \in P^-(\mathcal{B}, \boldsymbol{a}(\bar{\boldsymbol{x}})) \right\}$$

the cone of first order optimal gradients at $\bar{\boldsymbol{x}}$ w.r.t. $\boldsymbol{a}(\bar{\boldsymbol{x}}) \in \mathcal{B}$.

Theorem (Guignard's KKT Theorem)

Let H be closed and let G be some closed convex cone such that $K \cap G = P(\mathcal{F}, \bar{\boldsymbol{x}})$ and that $K^- + G^-$ is closed.

If $\bar{\boldsymbol{x}}$ minimizes $\psi(\boldsymbol{x})$ over $\mathcal{F}$, there is $\boldsymbol{\lambda} \in P^+(\mathcal{B}, \boldsymbol{a}(\bar{\boldsymbol{x}}))$ such that

$$-\nabla \psi(\bar{\boldsymbol{x}}) + \nabla \boldsymbol{a}(\bar{\boldsymbol{x}})) \cdot \boldsymbol{\lambda} \in G^-.$$

Proof (Guignard's KKT Theorem)

1. By the previous theorem, $-\nabla\psi(\bar{\boldsymbol{x}}) \in P^{-}(\mathcal{F}, \bar{\boldsymbol{x}})$.

Proof (Guignard's KKT Theorem)

1. By the previous theorem, $-\nabla\psi(\bar{\boldsymbol{x}}) \in P^{-}(\mathcal{F}, \bar{\boldsymbol{x}})$.
2. By assumption, $K^{-} + G^{-}$ is closed, hence $K^{-} + G^{-} = P^{-}(\mathcal{F}, \bar{\boldsymbol{x}})$.

Proof (Guignard's KKT Theorem)

1. By the previous theorem, $-\nabla\psi(\bar{\boldsymbol{x}}) \in P^{-}(\mathcal{F}, \bar{\boldsymbol{x}})$.
2. By assumption, $K^{-} + G^{-}$ is closed, hence $K^{-} + G^{-} = P^{-}(\mathcal{F}, \bar{\boldsymbol{x}})$.
3. Then, there is $\boldsymbol{q} \in K^{+}$ such that $-\nabla\psi(\bar{\boldsymbol{x}}) + \boldsymbol{q} \in G^{-}$.

Proof (Guignard's KKT Theorem)

1. By the previous theorem, $-\nabla\psi(\bar{\boldsymbol{x}}) \in P^{-}(\mathcal{F}, \bar{\boldsymbol{x}})$.
2. By assumption, $K^{-} + G^{-}$ is closed, hence $K^{-} + G^{-} = P^{-}(\mathcal{F}, \bar{\boldsymbol{x}})$.
3. Then, there is $\boldsymbol{q} \in K^{+}$ such that $-\nabla\psi(\bar{\boldsymbol{x}}) + \boldsymbol{q} \in G^{-}$.
4. Let $\boldsymbol{d} \in H^{-}$. Then $(\nabla \boldsymbol{a}(\bar{\boldsymbol{x}}) \cdot \boldsymbol{\lambda})^{T}\boldsymbol{d} \leqslant 0$ for all $\boldsymbol{\lambda} \in P(\mathcal{B}, \boldsymbol{a}(\bar{\boldsymbol{x}}))$.

Proof (Guignard's KKT Theorem)

1. By the previous theorem, $-\nabla\psi(\bar{\boldsymbol{x}}) \in P^-(\mathcal{F}, \bar{\boldsymbol{x}})$.
2. By assumption, $K^- + G^-$ is closed, hence $K^- + G^- = P^-(\mathcal{F}, \bar{\boldsymbol{x}})$.
3. Then, there is $\boldsymbol{q} \in K^+$ such that $-\nabla\psi(\bar{\boldsymbol{x}}) + \boldsymbol{q} \in G^-$.
4. Let $\boldsymbol{d} \in H^-$. Then $(\nabla\boldsymbol{a}(\bar{\boldsymbol{x}}) \cdot \boldsymbol{\lambda})^T\boldsymbol{d} \leqslant 0$ for all $\boldsymbol{\lambda} \in P(\mathcal{B}, \boldsymbol{a}(\bar{\boldsymbol{x}}))$.
5. Suppose now that $\nabla\boldsymbol{a}(\bar{\boldsymbol{x}})^T\boldsymbol{d} \notin P(\mathcal{B}, \boldsymbol{a}(\bar{\boldsymbol{x}}))$. Since $P(\mathcal{B}, \boldsymbol{a}(\bar{\boldsymbol{x}}))$ is a cone, separation ($\mathbb{R}^k$ is a separable Banach space) yields existence of an element $\boldsymbol{\mu} \in \mathbb{R}^m$ with

$$(\nabla\boldsymbol{a}(\bar{\boldsymbol{x}}) \cdot \boldsymbol{\mu})^T\boldsymbol{d} > 0 \geqslant \boldsymbol{\mu}^T\boldsymbol{\lambda} \quad \forall\boldsymbol{\lambda} \in P(\mathcal{B}, \boldsymbol{a}(\bar{\boldsymbol{x}})).$$

Proof (Guignard's KKT Theorem)

1. By the previous theorem, $-\nabla\psi(\bar{\boldsymbol{x}}) \in P^-(\mathcal{F}, \bar{\boldsymbol{x}})$.
2. By assumption, $K^- + G^-$ is closed, hence $K^- + G^- = P^-(\mathcal{F}, \bar{\boldsymbol{x}})$.
3. Then, there is $\boldsymbol{q} \in K^+$ such that $-\nabla\psi(\bar{\boldsymbol{x}}) + \boldsymbol{q} \in G^-$.
4. Let $\boldsymbol{d} \in H^-$. Then $(\nabla \boldsymbol{a}(\bar{\boldsymbol{x}}) \cdot \boldsymbol{\lambda})^T \boldsymbol{d} \leqslant 0$ for all $\boldsymbol{\lambda} \in P(\mathcal{B}, \boldsymbol{a}(\bar{\boldsymbol{x}}))$.
5. Suppose now that $\nabla \boldsymbol{a}(\bar{\boldsymbol{x}})^T \boldsymbol{d} \notin P(\mathcal{B}, \boldsymbol{a}(\bar{\boldsymbol{x}}))$. Since $P(\mathcal{B}, \boldsymbol{a}(\bar{\boldsymbol{x}}))$ is a cone, separation ($\mathbb{R}^k$ is a separable Banach space) yields existence of an element $\boldsymbol{\mu} \in \mathbb{R}^m$ with

$$(\nabla \boldsymbol{a}(\bar{\boldsymbol{x}}) \cdot \boldsymbol{\mu})^T \boldsymbol{d} > 0 \geqslant \boldsymbol{\mu}^T \boldsymbol{\lambda} \quad \forall \boldsymbol{\lambda} \in P(\mathcal{B}, \boldsymbol{a}(\bar{\boldsymbol{x}})).$$

6. Therefore, $\boldsymbol{\mu} \in P^-(\mathcal{B}, \boldsymbol{a}(\bar{\boldsymbol{x}}))$ and $\nabla \boldsymbol{a}(\bar{\boldsymbol{x}}) \cdot \boldsymbol{\mu} \in H$.

Proof (Guignard's KKT Theorem)

1. By the previous theorem, $-\nabla\psi(\bar{\boldsymbol{x}}) \in P^-(\mathcal{F}, \bar{\boldsymbol{x}})$.
2. By assumption, $K^- + G^-$ is closed, hence $K^- + G^- = P^-(\mathcal{F}, \bar{\boldsymbol{x}})$.
3. Then, there is $\boldsymbol{q} \in K^+$ such that $-\nabla\psi(\bar{\boldsymbol{x}}) + \boldsymbol{q} \in G^-$.
4. Let $\boldsymbol{d} \in H^-$. Then $(\nabla \boldsymbol{a}(\bar{\boldsymbol{x}}) \cdot \boldsymbol{\lambda})^T \boldsymbol{d} \leqslant 0$ for all $\boldsymbol{\lambda} \in P(\mathcal{B}, \boldsymbol{a}(\bar{\boldsymbol{x}}))$.
5. Suppose now that $\nabla \boldsymbol{a}(\bar{\boldsymbol{x}})^T \boldsymbol{d} \notin P(\mathcal{B}, \boldsymbol{a}(\bar{\boldsymbol{x}}))$. Since $P(\mathcal{B}, \boldsymbol{a}(\bar{\boldsymbol{x}}))$ is a cone, separation ($\mathbb{R}^k$ is a separable Banach space) yields existence of an element $\boldsymbol{\mu} \in \mathbb{R}^m$ with
$$(\nabla \boldsymbol{a}(\bar{\boldsymbol{x}}) \cdot \boldsymbol{\mu})^T \boldsymbol{d} > 0 \geqslant \boldsymbol{\mu}^T \boldsymbol{\lambda} \quad \forall \boldsymbol{\lambda} \in P(\mathcal{B}, \boldsymbol{a}(\bar{\boldsymbol{x}})).$$
6. Therefore, $\boldsymbol{\mu} \in P^-(\mathcal{B}, \boldsymbol{a}(\bar{\boldsymbol{x}}))$ and $\nabla \boldsymbol{a}(\bar{\boldsymbol{x}}) \cdot \boldsymbol{\mu} \in H$.
7. But this contradicts $(\nabla \boldsymbol{a}(\bar{\boldsymbol{x}}) \cdot \boldsymbol{\mu})^T \boldsymbol{d} > 0$.

Proof (Guignard's KKT Theorem)

1. By the previous theorem, $-\nabla\psi(\bar{\boldsymbol{x}}) \in P^{-}(\mathcal{F}, \bar{\boldsymbol{x}})$.
2. By assumption, $K^{-} + G^{-}$ is closed, hence $K^{-} + G^{-} = P^{-}(\mathcal{F}, \bar{\boldsymbol{x}})$.
3. Then, there is $\boldsymbol{q} \in K^{+}$ such that $-\nabla\psi(\bar{\boldsymbol{x}}) + \boldsymbol{q} \in G^{-}$.
4. Let $\boldsymbol{d} \in H^{-}$. Then $(\nabla\boldsymbol{a}(\bar{\boldsymbol{x}}) \cdot \boldsymbol{\lambda})^T\boldsymbol{d} \leqslant 0$ for all $\boldsymbol{\lambda} \in P(\mathcal{B}, \boldsymbol{a}(\bar{\boldsymbol{x}}))$.
5. Suppose now that $\nabla\boldsymbol{a}(\bar{\boldsymbol{x}})^T\boldsymbol{d} \notin P(\mathcal{B}, \boldsymbol{a}(\bar{\boldsymbol{x}}))$. Since $P(\mathcal{B}, \boldsymbol{a}(\bar{\boldsymbol{x}}))$ is a cone, separation ($\mathbb{R}^k$ is a separable Banach space) yields existence of an element $\boldsymbol{\mu} \in \mathbb{R}^m$ with

$$(\nabla\boldsymbol{a}(\bar{\boldsymbol{x}}) \cdot \boldsymbol{\mu})^T\boldsymbol{d} > 0 \geqslant \boldsymbol{\mu}^T\boldsymbol{\lambda} \quad \forall\boldsymbol{\lambda} \in P(\mathcal{B}, \boldsymbol{a}(\bar{\boldsymbol{x}})).$$

6. Therefore, $\boldsymbol{\mu} \in P^{-}(\mathcal{B}, \boldsymbol{a}(\bar{\boldsymbol{x}}))$ and $\nabla\boldsymbol{a}(\bar{\boldsymbol{x}}) \cdot \boldsymbol{\mu} \in H$.
7. But this contradicts $(\nabla\boldsymbol{a}(\bar{\boldsymbol{x}}) \cdot \boldsymbol{\mu})^T\boldsymbol{d} > 0$.
8. Hence $\nabla\boldsymbol{a}(\bar{\boldsymbol{x}})^T\boldsymbol{d} \in P(\mathcal{B}, \boldsymbol{a}(\bar{\boldsymbol{x}}))$.

Proof (Guignard's KKT Theorem)

1. By the previous theorem, $-\nabla\psi(\bar{\boldsymbol{x}}) \in P^{-}(\mathcal{F}, \bar{\boldsymbol{x}})$.
2. By assumption, $K^{-} + G^{-}$ is closed, hence $K^{-} + G^{-} = P^{-}(\mathcal{F}, \bar{\boldsymbol{x}})$.
3. Then, there is $\boldsymbol{q} \in K^{+}$ such that $-\nabla\psi(\bar{\boldsymbol{x}}) + \boldsymbol{q} \in G^{-}$.
4. Let $\boldsymbol{d} \in H^{-}$. Then $(\nabla \boldsymbol{a}(\bar{\boldsymbol{x}}) \cdot \boldsymbol{\lambda})^T \boldsymbol{d} \leqslant 0$ for all $\boldsymbol{\lambda} \in P(\mathcal{B}, \boldsymbol{a}(\bar{\boldsymbol{x}}))$.
5. Suppose now that $\nabla \boldsymbol{a}(\bar{\boldsymbol{x}})^T \boldsymbol{d} \notin P(\mathcal{B}, \boldsymbol{a}(\bar{\boldsymbol{x}}))$. Since $P(\mathcal{B}, \boldsymbol{a}(\bar{\boldsymbol{x}}))$ is a cone, separation ($\mathbb{R}^k$ is a separable Banach space) yields existence of an element $\boldsymbol{\mu} \in \mathbb{R}^m$ with

$$(\nabla \boldsymbol{a}(\bar{\boldsymbol{x}}) \cdot \boldsymbol{\mu})^T \boldsymbol{d} > 0 \geqslant \boldsymbol{\mu}^T \boldsymbol{\lambda} \quad \forall \boldsymbol{\lambda} \in P(\mathcal{B}, \boldsymbol{a}(\bar{\boldsymbol{x}})).$$

6. Therefore, $\boldsymbol{\mu} \in P^{-}(\mathcal{B}, \boldsymbol{a}(\bar{\boldsymbol{x}}))$ and $\nabla \boldsymbol{a}(\bar{\boldsymbol{x}}) \cdot \boldsymbol{\mu} \in H$.
7. But this contradicts $(\nabla \boldsymbol{a}(\bar{\boldsymbol{x}}) \cdot \boldsymbol{\mu})^T \boldsymbol{d} > 0$.
8. Hence $\nabla \boldsymbol{a}(\bar{\boldsymbol{x}})^T \boldsymbol{d} \in P(\mathcal{B}, \boldsymbol{a}(\bar{\boldsymbol{x}}))$.
9. That is, $\boldsymbol{d} \in K$ holds for all $\boldsymbol{d} \in H^{-}$. In other words, $H^{-} \subseteq K$.

Proof (Guignard's KKT Theorem)

1. By the previous theorem, $-\nabla\psi(\bar{\boldsymbol{x}}) \in P^-(\mathcal{F}, \bar{\boldsymbol{x}})$.
2. By assumption, $K^- + G^-$ is closed, hence $K^- + G^- = P^-(\mathcal{F}, \bar{\boldsymbol{x}})$.
3. Then, there is $\boldsymbol{q} \in K^+$ such that $-\nabla\psi(\bar{\boldsymbol{x}}) + \boldsymbol{q} \in G^-$.
4. Let $\boldsymbol{d} \in H^-$. Then $(\nabla \boldsymbol{a}(\bar{\boldsymbol{x}}) \cdot \boldsymbol{\lambda})^T \boldsymbol{d} \leqslant 0$ for all $\boldsymbol{\lambda} \in P(\mathcal{B}, \boldsymbol{a}(\bar{\boldsymbol{x}}))$.
5. Suppose now that $\nabla \boldsymbol{a}(\bar{\boldsymbol{x}})^T \boldsymbol{d} \notin P(\mathcal{B}, \boldsymbol{a}(\bar{\boldsymbol{x}}))$. Since $P(\mathcal{B}, \boldsymbol{a}(\bar{\boldsymbol{x}}))$ is a cone, separation ($\mathbb{R}^k$ is a separable Banach space) yields existence of an element $\boldsymbol{\mu} \in \mathbb{R}^m$ with
$$(\nabla \boldsymbol{a}(\bar{\boldsymbol{x}}) \cdot \boldsymbol{\mu})^T \boldsymbol{d} > 0 \geqslant \boldsymbol{\mu}^T \boldsymbol{\lambda} \quad \forall \boldsymbol{\lambda} \in P(\mathcal{B}, \boldsymbol{a}(\bar{\boldsymbol{x}})).$$
6. Therefore, $\boldsymbol{\mu} \in P^-(\mathcal{B}, \boldsymbol{a}(\bar{\boldsymbol{x}}))$ and $\nabla \boldsymbol{a}(\bar{\boldsymbol{x}}) \cdot \boldsymbol{\mu} \in H$.
7. But this contradicts $(\nabla \boldsymbol{a}(\bar{\boldsymbol{x}}) \cdot \boldsymbol{\mu})^T \boldsymbol{d} > 0$.
8. Hence $\nabla \boldsymbol{a}(\bar{\boldsymbol{x}})^T \boldsymbol{d} \in P(\mathcal{B}, \boldsymbol{a}(\bar{\boldsymbol{x}}))$.
9. That is, $\boldsymbol{d} \in K$ holds for all $\boldsymbol{d} \in H^-$. In other words, $H^- \subseteq K$.
10. Because H and K are closed convex cones, $K^- \subseteq H$ follows.

Proof (Guignard's KKT Theorem)

1. By the previous theorem, $-\nabla\psi(\bar{\boldsymbol{x}}) \in P^-(\mathcal{F}, \bar{\boldsymbol{x}})$.
2. By assumption, $K^- + G^-$ is closed, hence $K^- + G^- = P^-(\mathcal{F}, \bar{\boldsymbol{x}})$.
3. Then, there is $\boldsymbol{q} \in K^+$ such that $-\nabla\psi(\bar{\boldsymbol{x}}) + \boldsymbol{q} \in G^-$.
4. Let $\boldsymbol{d} \in H^-$. Then $(\nabla\boldsymbol{a}(\bar{\boldsymbol{x}}) \cdot \boldsymbol{\lambda})^T\boldsymbol{d} \leqslant 0$ for all $\boldsymbol{\lambda} \in P(\mathcal{B}, \boldsymbol{a}(\bar{\boldsymbol{x}}))$.
5. Suppose now that $\nabla\boldsymbol{a}(\bar{\boldsymbol{x}})^T\boldsymbol{d} \notin P(\mathcal{B}, \boldsymbol{a}(\bar{\boldsymbol{x}}))$. Since $P(\mathcal{B}, \boldsymbol{a}(\bar{\boldsymbol{x}}))$ is a cone, separation ($\mathbb{R}^k$ is a separable Banach space) yields existence of an element $\boldsymbol{\mu} \in \mathbb{R}^m$ with
$$(\nabla\boldsymbol{a}(\bar{\boldsymbol{x}}) \cdot \boldsymbol{\mu})^T\boldsymbol{d} > 0 \geqslant \boldsymbol{\mu}^T\boldsymbol{\lambda} \quad \forall\boldsymbol{\lambda} \in P(\mathcal{B}, \boldsymbol{a}(\bar{\boldsymbol{x}})).$$
6. Therefore, $\boldsymbol{\mu} \in P^-(\mathcal{B}, \boldsymbol{a}(\bar{\boldsymbol{x}}))$ and $\nabla\boldsymbol{a}(\bar{\boldsymbol{x}}) \cdot \boldsymbol{\mu} \in H$.
7. But this contradicts $(\nabla\boldsymbol{a}(\bar{\boldsymbol{x}}) \cdot \boldsymbol{\mu})^T\boldsymbol{d} > 0$.
8. Hence $\nabla\boldsymbol{a}(\bar{\boldsymbol{x}})^T\boldsymbol{d} \in P(\mathcal{B}, \boldsymbol{a}(\bar{\boldsymbol{x}}))$.
9. That is, $\boldsymbol{d} \in K$ holds for all $\boldsymbol{d} \in H^-$. In other words, $H^- \subseteq K$.
10. Because H and K are closed convex cones, $K^- \subseteq H$ follows.
11. Then, there is $\boldsymbol{\lambda} \in P^+(\mathcal{B}, \boldsymbol{a}(\bar{\boldsymbol{x}}))$ with $-\nabla\psi(\bar{\boldsymbol{x}}) + \nabla\boldsymbol{a}(\bar{\boldsymbol{x}}) \cdot \boldsymbol{\lambda} \in G^-$. $\square$

Application to NLPs

Let us reconcile this with the KKT theorem for NLPs you all know.

$$\boxed{\begin{array}{rl} \displaystyle\min_{\boldsymbol{x}\in\mathbb{R}^n,\boldsymbol{s}\in\mathbb{R}^c} & F(x) \\ \text{s.t.} & G(\boldsymbol{x}) = 0,\ H(\boldsymbol{x}) - \boldsymbol{s} = 0 \\ & \boldsymbol{s} \geqslant \boldsymbol{0} \end{array}}$$

In our setting $\mathcal{F} = \{(\boldsymbol{x}, \boldsymbol{s}) \in \mathcal{C} \mid \boldsymbol{a}(\boldsymbol{x}, \boldsymbol{s}) \in \mathcal{B}\}$, we have

$$\boldsymbol{a}(\boldsymbol{x}) = (G(\boldsymbol{x}),\ H(\boldsymbol{x}) - \boldsymbol{s}),\ \mathcal{B} = \{\boldsymbol{0}\},\ \mathcal{C} = \mathbb{R}^n \times \mathbb{R}^c_{\geqslant 0}.$$

Cones:

- $P(\mathcal{B}, \boldsymbol{a}(\bar{\boldsymbol{x}}, \bar{\boldsymbol{s}})) = \{\boldsymbol{0}\}$
- $K = \{(\boldsymbol{d}_x, \boldsymbol{d}_s) \mid \nabla G(\bar{\boldsymbol{x}})^T \boldsymbol{d}_x = \boldsymbol{0},\ \nabla H(\bar{\boldsymbol{x}})^T \boldsymbol{d}_x - \boldsymbol{d}_s = \boldsymbol{0}\}$
- $H = \{(\boldsymbol{q}_x, \boldsymbol{q}_s) \mid \boldsymbol{q}_x = \nabla G(\bar{\boldsymbol{x}}) \cdot \boldsymbol{\lambda}_G + \nabla H(\bar{\boldsymbol{x}}) \cdot \boldsymbol{\lambda}_H,\ \boldsymbol{q}_s = -\boldsymbol{\lambda}_H,\ (\boldsymbol{\lambda}_G, \boldsymbol{\lambda}_H) \in \mathbb{R}^{2c}\}$

Also, K and H are **closed convex cones**, as required.

Guignard's Constraint Qualification (GCQ)

- We're supposed to choose a set G to satisfy the prerequisite. If we let

$$G = P(\mathcal{C}, \bar{\boldsymbol{x}}, \bar{\boldsymbol{s}}) = \{(\boldsymbol{d}_x, \boldsymbol{d}_s) \in \mathbb{R}^{n+c} \mid d_{s,i} \geqslant 0 \text{ if } \bar{s}_i = 0\},$$

  the prerequisite of the theorem $K \cap G \overset{!}{=} P(\mathcal{F}, \bar{\boldsymbol{x}})$ reads

$$\begin{aligned} &\{(\boldsymbol{d}_x, \boldsymbol{d}_s) \mid \nabla G(\bar{\boldsymbol{x}})^T \boldsymbol{d}_x = \mathbf{0},\ \nabla H(\bar{\boldsymbol{x}})^T \boldsymbol{d}_x - \boldsymbol{d}_s = \mathbf{0}, \boldsymbol{d}_{s,i} \leqslant \mathbf{0} \text{ if } s_i = 0\} \\ \overset{!}{=}\ & P(\mathcal{F}, \bar{\boldsymbol{x}}) = \overline{\operatorname{conv} T(\mathcal{F}, \bar{\boldsymbol{x}})} = (\operatorname{conv} T(\mathcal{F}, \bar{\boldsymbol{x}}))^{--}. \end{aligned}$$

- The left hand side is the linearized cone, and the prerequisite simplifies to

$$L(\mathcal{F}, \bar{\boldsymbol{x}})^- = T(\mathcal{F}, \bar{\boldsymbol{x}})^-,$$

  "the dual of the linearized cone must equal the dual of the tangent cone".

- This is called **Guignard's Constraint Qualification (GCQ)**, and is the weakest condition under which a variant of the KKT theorem can be proven.

KKT Theorem under GCQ

- The theorem's statement now reads:

  If $\bar{\boldsymbol{x}}$ minimizes $\psi(\boldsymbol{x})$ over $\mathcal{F}$ and GCQ holds for $\mathcal{F}$ at $\bar{\boldsymbol{x}}$, there are $\boldsymbol{\lambda}_G, \boldsymbol{\lambda}_H$ such that

$$\begin{aligned} -\nabla_x \psi(\bar{\boldsymbol{x}}) + \nabla G(\bar{\boldsymbol{x}}) \cdot \boldsymbol{\lambda}_G + \nabla H(\bar{\boldsymbol{x}}) \cdot \boldsymbol{\lambda}_H &= 0 \\ \lambda_{H,i} &= 0 \text{ if } s_i > 0 \\ \lambda_{H,i} &\leqslant 0 \text{ if } s_i = 0 \end{aligned}$$

  This is the usual form of the KKT theorem.

- Notes:
  $P^+(\mathcal{B}, \boldsymbol{a}(\bar{\boldsymbol{x}})) = \mathbb{R}^c$
  $G^-(\mathcal{F}, \boldsymbol{x}, \boldsymbol{s}) = \{0\}^n \times \{\lambda_H \mid \lambda_{H,i} = 0 \text{ if } s_i > 0,\ \lambda_{H,i} \leqslant 0 \text{ if } s_i = 0\}$

Application to MPCCs: First Attempt

Let's now try to obtain an optimality condition for the most simple MPCC

$$\boxed{\min \psi(u, v) \text{ s.t. } 0 \leqslant u \perp v \geqslant 0.}$$

Certainly, ψ can be chosen such that $\boldsymbol{x}^* = (u^*, v^*) = (0, 0)$ is a minimizer.

Encode $\mathcal{F}$: $\boldsymbol{a}(\boldsymbol{x}) = u \cdot v$, $\mathcal{B} = \{0\}$, $\mathcal{C} = \mathbb{R}^2_{\geqslant 0}$.

Cones:

- $K = \mathbb{R}^2$ because $\nabla \boldsymbol{a}(\bar{\boldsymbol{x}}) = \mathbf{0}$, and $P(\mathcal{B}, \boldsymbol{a}(\bar{\boldsymbol{x}})) = \{\mathbf{0}\}$, $P^+(\mathcal{B}, \boldsymbol{a}(\bar{\boldsymbol{x}})) = \mathbb{R}^2$
- $P(\mathcal{F}, \bar{\boldsymbol{x}}) = \mathcal{F}$, a cone but not convex; $P^-(\mathcal{F}, \bar{\boldsymbol{x}}) = \mathbb{R}^2_{\leqslant 0}$
- $G = P(\mathcal{F}, \bar{\boldsymbol{x}})$ is the only (and trivial) choice that satisfies $K \cap G = \mathcal{P}(\mathcal{F}, \bar{\boldsymbol{x}})$
- Then $G^- = \mathbb{R}^2_{\leqslant 0}$

Guignard's KKT theorem now yields the following statement:

If $\bar{\boldsymbol{x}}$ minimizes ψ over $\mathcal{F}$,
"there is $\boldsymbol{\lambda} \in \mathbb{R}^2$ such that $-\nabla \psi(\bar{\boldsymbol{x}}) + \nabla \boldsymbol{a}(\bar{\boldsymbol{x}}) \cdot \boldsymbol{\lambda} \leqslant \mathbf{0}$".

Application to MPCCs: Second Attempt

This is very weak. Clearly, our encoding is to blame. Here's a better idea:

Encode: $\boldsymbol{a}(\boldsymbol{x}) = (u, v)$, $\mathcal{B} = (\mathbb{R}_{\geqslant 0} \times \{0\}) \cup (\{0\} \times \mathbb{R}_{\geqslant 0})$, $\mathcal{C} = \mathbb{R}^2$.

Cones:

- $K = \mathcal{B}$ because $\nabla \boldsymbol{a}(\bar{\boldsymbol{x}}) = \boldsymbol{I}$, and $P(\mathcal{B}, \boldsymbol{a}(\bar{\boldsymbol{x}})) = \mathcal{B}$, $P^+(\mathcal{B}, \boldsymbol{a}(\bar{\boldsymbol{x}})) = \mathbb{R}^2_{\geqslant 0}$
- $G = \mathbb{R}^2$ now satisfies $K \cap G = \mathcal{P}(\mathcal{F}, \bar{\boldsymbol{x}})$
- Then $G^- = \{\boldsymbol{0}\}$

Guignard's theorem now yields the following, much improved statement:

If $\bar{\boldsymbol{x}}$ minimizes ψ over $\mathcal{F}$,
"there is $\boldsymbol{\lambda} \in \mathbb{R}^2_{\geqslant 0}$ such that $-\nabla\psi(\bar{\boldsymbol{x}}) + \nabla \boldsymbol{a}(\bar{\boldsymbol{x}}) \cdot \boldsymbol{\lambda} = \boldsymbol{0}$."

MPCC-Lagrangian Function

The second encoding **discourages** the notation $u \cdot v = 0$ and **motivates** the introduction of a pair of multiplier vectors for the complementarity constraint:

MPCC as NLP:

$$\boxed{\begin{aligned} &\min_{\boldsymbol{x}\in\mathbb{R}^n} F(\boldsymbol{x}) \\ &\text{ s.t. } \mathbf{0} = G(\boldsymbol{x}) \cdot H(\boldsymbol{x}) \\ &\qquad\ \mathbf{0} \leqslant G(\boldsymbol{x}), \mathbf{0} \leqslant H(\boldsymbol{x}) \end{aligned}}$$

NLP-Lagrangian:

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}_{GH}, \boldsymbol{\mu}_G, \boldsymbol{\mu}_H) := F(\boldsymbol{x}) - \boldsymbol{\lambda}_{GH}^T(G(x) \cdot H(\boldsymbol{x})) - \boldsymbol{\mu}_G^T G(\boldsymbol{x}) - \boldsymbol{\mu}_H^T H(\boldsymbol{x}).$$

MPCC:

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

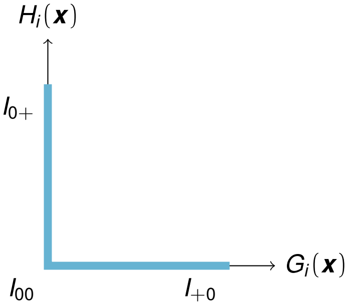
MPCC-Lagrangian:

$$\mathcal{L}_{\text{MPCC}}(\boldsymbol{x}, \boldsymbol{\mu}_G, \boldsymbol{\mu}_H) := F(\boldsymbol{x}) - \boldsymbol{\mu}_G^T G(\boldsymbol{x}) - \boldsymbol{\mu}_H^T H(\boldsymbol{x})$$

Active Sets

Active sets of **strict** complementarities:
$I_{+0}(\bar{\boldsymbol{x}}) := \{i \mid G_i(\bar{\boldsymbol{x}}) > 0,\ H_i(\bar{\boldsymbol{x}}) = 0\}$
$I_{0+}(\bar{\boldsymbol{x}}) := \{i \mid G_i(\bar{\boldsymbol{x}}) = 0,\ H_i(\bar{\boldsymbol{x}}) > 0\}$

Active set of **non-strict** complementarity pairs (**biactive set**):
$I_{00}(\bar{\boldsymbol{x}}) := \{i \mid G_i(\bar{\boldsymbol{x}}) = 0,\ H_i(\bar{\boldsymbol{x}}) = 0\}$



We say that **Lower-Level Strict Complementarity (LLSCC)** is satisfied at $\bar{\boldsymbol{x}}$ if $I_{00}(\bar{\boldsymbol{x}}) = \emptyset$. Then, the constraint $\mathbf{0} = G(\boldsymbol{x}) \cdot H(\boldsymbol{x})$ can locally be disposed of. The MPCC locally looks like an NLP satisfying constraint qualifications.

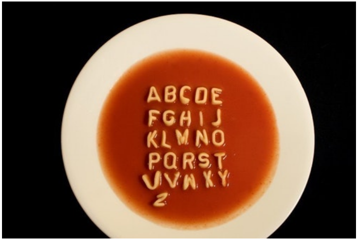
Assuming (LLSCC) is usually held for way too strong a restriction to be of practical interest.

Strong Stationarity (S)

Remember that GCQ had a chance of being satisfied by an MPCC.

Then, if $\bar{\boldsymbol{x}}$ minimizes ψ over $\mathcal{F}$, there is $\boldsymbol{\lambda} \in P^{+}(\mathcal{B}, \boldsymbol{a}(\bar{\boldsymbol{x}}))$ such that

$$-\nabla\psi(\bar{\boldsymbol{x}}) + \nabla\boldsymbol{a}(\bar{\boldsymbol{x}}) \cdot \boldsymbol{\lambda} \in P^{-}(\mathcal{F}, \bar{\boldsymbol{x}}).$$



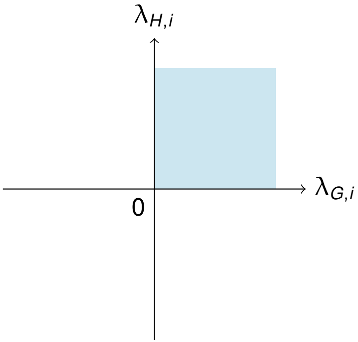
Alphabet soup of stationarity conditions.

Strong or S-Stationarity: If $\bar{\boldsymbol{x}} \in \mathcal{F}$ is a local minimizer of MPCC and GCQ holds at $\bar{\boldsymbol{x}}$, there are multipliers $\boldsymbol{\lambda}_C$, $\boldsymbol{\mu}_D$, $\boldsymbol{\mu}_G$, $\boldsymbol{\mu}_H$ such that

$$\nabla\mathcal{L}_{\text{MPCC}}(\bar{\boldsymbol{x}}, \boldsymbol{\lambda}_C, \boldsymbol{\mu}_D, \boldsymbol{\mu}_G, \boldsymbol{\mu}_H) = \mathbf{0}$$

$$\begin{aligned} \boldsymbol{\mu}_D &\geqslant \mathbf{0} \\ \mu_{D,i} &= 0 \text{ if } D_i(\bar{\boldsymbol{x}}) > 0 \\ \mu_{G,i} \geqslant 0,\ \mu_{H,i} &\geqslant 0 \text{ if } i \in I_{00}(\bar{\boldsymbol{x}}) \\ \mu_{H,i} &= 0 \text{ if } i \in I_{0+}(\bar{\boldsymbol{x}}) \\ \mu_{G,i} &= 0 \text{ if } i \in I_{+0}(\bar{\boldsymbol{x}}) \end{aligned}$$

S-Stationarity multiplier set



Optimal multiplier signs for $i \in I_{00}(\bar{\boldsymbol{x}})$ under S-stationarity.

MPCC-Linear Independence CQ

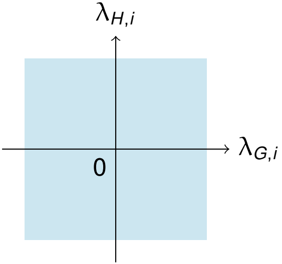
We say that **MPCC-Linear Independence CQ (MPCC-LICQ)** holds at $\bar{\boldsymbol{x}} \in \mathcal{F}$ if the gradients

$$\begin{pmatrix} \nabla C(\bar{\boldsymbol{x}}) & \nabla D_i(\bar{\boldsymbol{x}}) & \nabla G_i(\bar{\boldsymbol{x}}) & \nabla H_i(\bar{\boldsymbol{x}}) \end{pmatrix}$$
$$\qquad D_i(\bar{\boldsymbol{x}}) = 0 \quad i \in I_{0+}(\bar{\boldsymbol{x}}) \cup I_{00}(\bar{\boldsymbol{x}}) \quad i \in I_{+0}(\bar{\boldsymbol{x}}) \cup I_{00}(\bar{\boldsymbol{x}})$$

are linearly independent.

MPCC-LICQ is LICQ for the **tightened NLP** at $\bar{\boldsymbol{x}}$:

$$\begin{aligned} \min_{\boldsymbol{x} \in \mathbb{R}^n} \ & F(\boldsymbol{x}) \\ \text{s.t. } & C(\boldsymbol{x}) = \mathbf{0} \\ & D(\boldsymbol{x}) \geqslant \mathbf{0} \\ & G_i(\boldsymbol{x}) = 0,\ H_i(\boldsymbol{x}) \geqslant 0 \text{ if } i \in I_{0+}(\bar{\boldsymbol{x}}) \\ & G_i(\boldsymbol{x}) \geqslant 0,\ H_i(\boldsymbol{x}) = 0 \text{ if } i \in I_{+0}(\bar{\boldsymbol{x}}) \\ & G_i(\boldsymbol{x}) = 0,\ H_i(\boldsymbol{x}) = 0 \text{ if } i \in I_{00}(\bar{\boldsymbol{x}}) \end{aligned}$$



Optimal multiplier signs for TNLP($\bar{\boldsymbol{x}}$) in $I_{00}(\bar{\boldsymbol{x}})$.

MPCC-LICQ implies GCQ

Theorem

If MPCC-LICQ holds at $\bar{\boldsymbol{x}} \in \mathcal{F}$, then GCQ holds at $\bar{\boldsymbol{x}}$.

Sketch of proof: For sets $P \subseteq I_{00}(\bar{\boldsymbol{x}})$ we have

$$T(\mathcal{F}, \bar{\boldsymbol{x}}) = \bigcup_{P \subseteq I_{00}(\bar{\boldsymbol{x}})} T(\mathcal{F}(P), \bar{\boldsymbol{x}}) \quad \Longrightarrow \quad T(\mathcal{F}, \bar{\boldsymbol{x}})^\circ = \bigcap_{P \subseteq I_{00}(\bar{\boldsymbol{x}})} T(\mathcal{F}(P), \bar{\boldsymbol{x}})^\circ.$$

If a constraint qualification holds at $\bar{\boldsymbol{x}}$ for all sets $\mathcal{F}(P)$, we have

$$T(\mathcal{F}, \bar{\boldsymbol{x}})^\circ = \bigcap_{P \subseteq I_{00}(\bar{\boldsymbol{x}})} L(\mathcal{F}(P), \bar{\boldsymbol{x}})^\circ.$$

For GCQ to hold at $\bar{\boldsymbol{x}}$, we now have to show $T(\mathcal{F}, \bar{\boldsymbol{x}})^\circ \subseteq L(\mathcal{F}, \bar{\boldsymbol{x}})^\circ$, as $T(\mathcal{F}, \bar{\boldsymbol{x}})^\circ \supseteq L(\mathcal{F}, \bar{\boldsymbol{x}})^\circ$ always holds.

MPCC-LICQ implies GCQ

Sketch of proof, continued: The cones $L(\mathcal{F}(P), \bar{\boldsymbol{x}})^\circ$ have the explicit representations $L(\mathcal{F}(P), \bar{\boldsymbol{x}}) =$

$$\left\{ \boldsymbol{w} \;\middle|\; \begin{array}{l} \boldsymbol{w} = \nabla C(\bar{\boldsymbol{x}}) \cdot \boldsymbol{\lambda}_C + \nabla D(\bar{\boldsymbol{x}}) \cdot \mu_D + \nabla G(\bar{\boldsymbol{x}}) \cdot \boldsymbol{\mu}_G + \nabla H(\bar{\boldsymbol{x}}) \cdot \boldsymbol{\mu}_H \\ \mu_{D,i} = 0 \text{ if } D_i(\bar{\boldsymbol{x}}) > 0,\ \mu_{D,i} \geqslant 0 \text{ if } D_i(\bar{\boldsymbol{x}}) \geqslant 0 \\ \mu_{G,i} = 0 \text{ if } i \in I_{+0}(\bar{\boldsymbol{x}}),\ \mu_{G,i} \geqslant 0 \text{ if } i \in P^{\mathrm{C}} \\ \mu_{H,i} = 0 \text{ if } i \in I_{0+}(\bar{\boldsymbol{x}}),\ \mu_{H,i} \geqslant 0 \text{ if } i \in P \end{array} \right\}.$$

MPCC-LICQ implies a CQ for every $\mathcal{F}(P)$. Now for any element $\boldsymbol{w} \in T(\mathcal{F}(P), \bar{\boldsymbol{x}})^\circ$, use this to show that for every P there exist vectors $\boldsymbol{\lambda}_C$, $\boldsymbol{\mu}_D$, $\boldsymbol{\mu}_G$, $\boldsymbol{\mu}_H$ such that $\boldsymbol{w} \in L(\mathcal{F}(P), \bar{\boldsymbol{x}})^\circ$.

Finally, MPCC-LICQ for all $\mathcal{F}(P)$ implies uniqueness of the multipliers across all $P \subset I_{00}(\bar{\boldsymbol{x}})$. Then $\boldsymbol{w} \in T(\mathcal{F}, \bar{\boldsymbol{x}})^\circ$ implies $\boldsymbol{w} \in L(\mathcal{F}, \bar{\boldsymbol{x}})^\circ$, which shows that GCQ holds. $\square$

S-Stationarity and Local Minimizers

Summarizing, we have just proven the following:

Theorem (S-Stationarity is necessary under MPCC-LICQ)

If MPCC-LICQ holds at $\bar{\boldsymbol{x}} \in \mathcal{F}$, then GCQ holds at $\bar{\boldsymbol{x}}$. If $\bar{\boldsymbol{x}}$ is a local minimum of MPCC, and GCQ holds at $\bar{\boldsymbol{x}}$, then $\bar{\boldsymbol{x}}$ is an S-stationary point.

“Convex” MPCCs satisfy GCQ and permit to also prove the converse:

Theorem (S-Stationarity is sufficient under MPCC-convexity)

If F is convex, D is concave, and C, G, and H are affine linear, then every S-stationary point of MPCC is a local minimum.

On the other hand, we have already seen that GCQ may not hold. Hence, MPCC-LICQ is usually considered too restrictive to serve as a working basis.

Non-smooth minimization

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

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

Example:
$\phi(u, v) = \min\{u, v\}$

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

Differentiable NCP-functions necessarily satisfy $\nabla\phi(0, 0) = (0, 0)^T$.

Useful NCP-functions are **nondifferentiable** in $(0, 0)$.

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:

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

Branch or Piece NLPs

We may pick an element of the Bouligand subdifferential by specifying a subset $P \subseteq I_{00}(\bar{\boldsymbol{x}})$ and its complement $P^{\mathrm{C}} := I_{00}(\bar{\boldsymbol{x}}) \setminus P$ relative to $I_{00}(\bar{\boldsymbol{x}})$.

Then define the **branch NLP** for P at $\bar{\boldsymbol{x}}$:

$$
(\mathrm{NLP}(\bar{\boldsymbol{x}}, P)) \qquad \boxed{\begin{aligned}
\min_{\boldsymbol{x} \in \mathbb{R}^n} \; & F(\boldsymbol{x}) \\
\text{s.t. } & C(\boldsymbol{x}) = \mathbf{0} \\
& D(\boldsymbol{x}) \geqslant \mathbf{0} \\
& G_i(\boldsymbol{x}) = 0, \; H_i(\boldsymbol{x}) \geqslant 0 \text{ if } i \in I_{0+}(\bar{\boldsymbol{x}}) \cup P \\
& G_i(\boldsymbol{x}) \geqslant 0, \; H_i(\boldsymbol{x}) = 0 \text{ if } i \in I_{+0}(\bar{\boldsymbol{x}}) \cup P^{\mathrm{C}}
\end{aligned}}
$$

There are $2^{|I_{00}(\bar{\boldsymbol{x}})|}$ branch NLPs in a point $\bar{\boldsymbol{x}} \in \mathcal{F}$. If LLSCC holds, there is only one. Then, in a small neighborhood of $\bar{\boldsymbol{x}}$, MPCC looks like that branch NLP.

Bouligand Stationarity (B)

Bouligand- or B-Stationarity: A point $\bar{\boldsymbol{x}} \in \mathcal{F}$ is called B-stationary if for every $P \subseteq I_{00}(\bar{\boldsymbol{x}})$ there are multipliers $\boldsymbol{\lambda}_C$, $\boldsymbol{\mu}_D$, $\boldsymbol{\mu}_G$, $\boldsymbol{\mu}_H$ (possibly depending on P) such that

$$\nabla \mathcal{L}_{\text{MPCC}}(\bar{\boldsymbol{x}}, \boldsymbol{\lambda}_C, \boldsymbol{\mu}_D, \boldsymbol{\mu}_G, \boldsymbol{\mu}_H) = \mathbf{0}$$

$$\begin{aligned}
\boldsymbol{\mu}_D &\geqslant \mathbf{0} \\
\mu_{D,i} &= 0 \text{ if } D_i(\bar{\boldsymbol{x}}) > 0 \\
\mu_{H,i} &= 0 \text{ if } i \in I_{0+}(\bar{\boldsymbol{x}}) \\
\mu_{G,i} &= 0 \text{ if } i \in I_{+0}(\bar{\boldsymbol{x}}) \\
\mu_{H,i} &\geqslant 0 \text{ if } i \in P \subset I_{00}(\bar{\boldsymbol{x}}) \\
\mu_{G,i} &\geqslant 0 \text{ if } i \in P^{\mathrm{C}} \subset I_{00}(\bar{\boldsymbol{x}})
\end{aligned}$$

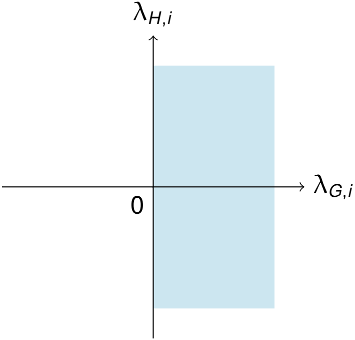
Theorem

A local minimizer $\bar{\boldsymbol{x}} \in \mathcal{F}$ of MPCC is B-stationary.

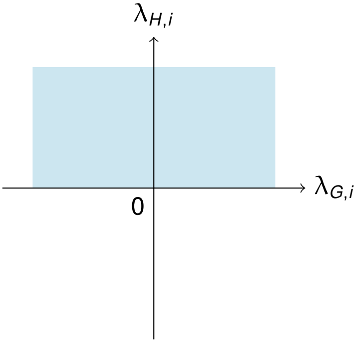
A piece $P \subset I_{00}(\bar{\boldsymbol{x}})$ with a non-optimal multiplier is a poly-size certificate for non-B-stationarity. The B-stationarity decision problem is in co-NP (in absence for further CQs) because there are $2^{|I_{00}(\bar{\boldsymbol{x}})|}$ pieces P to check.

B-stationarity multiplier sets

$$c = 1,\ I_{00}(\bar{\boldsymbol{x}}) = \{1\}.$$



Optimal multiplier signs for piece $P = \emptyset$.



Optimal multiplier signs for piece $P = \{1\}$.

Clarke subdifferential

The set

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

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:

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

Chain Rule for ∂^{C} :

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

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 $\bar{\boldsymbol{x}}$.

Using the Clarke subdifferential in KKT

Applying this chain rule to the NCP function $\phi_i(\bar{\boldsymbol{x}}) = \min(G_i(\bar{\boldsymbol{x}}), H_i(\bar{\boldsymbol{x}}))$ yields the estimate

$$\begin{aligned}\partial^{\mathrm{C}}\phi_i(\bar{\boldsymbol{x}}) &\subseteq \operatorname{conv}\{\nabla G_i(\bar{\boldsymbol{x}}), \nabla H_i(\bar{\boldsymbol{x}})\} \\ &= \Big\{(\xi_i \nabla G_i(\bar{\boldsymbol{x}}), (1-\xi_i)\nabla H_i(\bar{\boldsymbol{x}})) \;\Big|\; 0 \leqslant \xi_i \leqslant 1\Big\}\end{aligned}$$

and equality can be shown by a refined argument.

Using the Clarke subdifferential in KKT

Applying this chain rule to the NCP function $\phi_i(\bar{\boldsymbol{x}}) = \min(G_i(\bar{\boldsymbol{x}}), H_i(\bar{\boldsymbol{x}}))$ yields the estimate

$$\begin{aligned}\partial^{\mathrm{C}}\phi_i(\bar{\boldsymbol{x}}) &\subseteq \operatorname{conv}\{\nabla G_i(\bar{\boldsymbol{x}}), \nabla H_i(\bar{\boldsymbol{x}})\} \\ &= \Big\{(\xi_i \nabla G_i(\bar{\boldsymbol{x}}), (1-\xi_i)\nabla H_i(\bar{\boldsymbol{x}})) \;\Big|\; 0 \leqslant \xi_i \leqslant 1\Big\}\end{aligned}$$

and equality can be shown by a refined argument.

Inserting any particular element from $\partial^{\mathrm{C}}\phi_i(\bar{\boldsymbol{x}})$ (given by a $\boldsymbol{\xi} \in [0,1]^c$) into the KKT conditions yields

$$\nabla F(\bar{\boldsymbol{x}}) \in \sum_{i \in I_{0+}(\bar{\boldsymbol{x}})} \nabla G_i(\bar{\boldsymbol{x}}) \cdot \delta_i + \sum_{i \in I_{+0}(\bar{\boldsymbol{x}})} \nabla H_i(\bar{\boldsymbol{x}}) \cdot \delta_i + \sum_{i \in I_{00}(\bar{\boldsymbol{x}})} \operatorname{conv}\{\nabla G_i(\bar{\boldsymbol{x}}), \nabla H_i(\bar{\boldsymbol{x}})\} \cdot \delta_i$$

with MPCC multipliers

$$\lambda_{G,i} = \begin{cases} \delta_i & \text{if } i \in I_{0+}(\bar{\boldsymbol{x}}) \\ \xi_i \delta_i & \text{if } i \in I_{00}(\bar{\boldsymbol{x}}) \\ 0 & \text{if } i \in I_{+0}(\bar{\boldsymbol{x}}) \end{cases} \quad \text{and} \quad \lambda_{H,i} = \begin{cases} 0 & \text{if } i \in I_{0+}(\bar{\boldsymbol{x}}) \\ (1-\xi_i)\delta_i & \text{if } i \in I_{00}(\bar{\boldsymbol{x}}) \\ \delta_i & \text{if } i \in I_{+0}(\bar{\boldsymbol{x}}). \end{cases}$$

Clarke stationarity (C)

We may simplify the conditions on $\boldsymbol{\lambda}_G$ and $\boldsymbol{\lambda}_H$ for the biactive set to

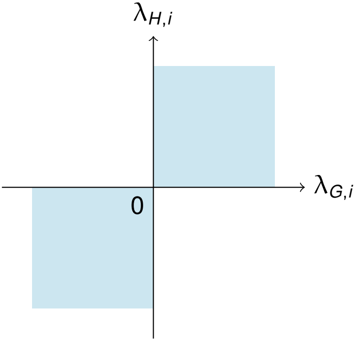
$$\lambda_{G,i} \cdot \lambda_{H,i} = \xi_i(1-\xi_i)\delta_i^2 \geqslant 0,\ \ i \in I_{00}(\bar{\boldsymbol{x}}).$$

A point $\boldsymbol{x} \in \mathcal{F}$ is called **Clarke-** or **C-stationarity** if there are multipliers $\boldsymbol{\lambda}_C$, $\boldsymbol{\mu}_D$, $\boldsymbol{\mu}_G$, $\boldsymbol{\mu}_H$ such that

$$\nabla\mathcal{L}_{\mathrm{MPCC}}(\bar{\boldsymbol{x}}, \boldsymbol{\lambda}_C, \boldsymbol{\mu}_D, \boldsymbol{\mu}_G, \boldsymbol{\mu}_H) = \mathbf{0}$$

$$\begin{aligned}
\boldsymbol{\mu}_D &\geqslant \mathbf{0}\\
\mu_{D,i} &= 0 \text{ if } D_i(\bar{\boldsymbol{x}}) > 0\\
\mu_{G,i} \cdot \mu_{H,i} &\geqslant 0 \text{ if } i \in I_{00}(\bar{\boldsymbol{x}})\\
\mu_{H,i} &= 0 \text{ if } i \in I_{0+}(\bar{\boldsymbol{x}})\\
\mu_{G,i} &= 0 \text{ if } i \in I_{+0}(\bar{\boldsymbol{x}})
\end{aligned}$$

C-stationarity multiplier set



Optimal multiplier signs for C-stationarity.

C-stationarity ignores trivial descent directions. Algorithmically, it is an unsatisfying concept for MPCCs.

Clarke stationarity (C)

MPCC-Mangasarian-Fromovitz CQ holds at $\bar{\boldsymbol{x}} \in \mathcal{F}$ if the gradients $\nabla C(\bar{\boldsymbol{x}})$ and $\nabla D_i(\bar{\boldsymbol{x}})$ for $D_i(\bar{\boldsymbol{x}}) = 0$, $\nabla G_i(\bar{\boldsymbol{x}})$ for $i \in I_{0+}(\bar{\boldsymbol{x}}) \cup I_{00}(\bar{\boldsymbol{x}})$, and $\nabla H_i(\bar{\boldsymbol{x}})$ for $i \in I_{+0}(\bar{\boldsymbol{x}}) \cup I_{00}(\bar{\boldsymbol{x}})$ are **positively linearly independent.**

This means that there are nontrivial multipliers $\boldsymbol{\lambda}_G$, $\boldsymbol{\mu}_D \geqslant \mathbf{0}$, $\boldsymbol{\mu}_G \geqslant \mathbf{0}$, and $\boldsymbol{\mu}_H \geqslant \mathbf{0}$ such that

$$\mathbf{0} = \nabla C(\bar{\boldsymbol{x}}) \cdot \boldsymbol{\lambda} + \sum_{i:\ D_i \geqslant 0} \nabla D_i(\bar{\boldsymbol{x}}) \cdot \mu_{D,i} + \sum_{I_{0+} \cup I_{00}} \nabla G_i(\bar{\boldsymbol{x}}) \cdot \mu_{G,i} + \sum_{I_{+0} \cup I_{00}} \nabla H_i(\bar{\boldsymbol{x}}) \cdot \mu_{H,i}.$$

MPCC-LICQ at $\bar{\boldsymbol{x}}$ implies MPCC-MFCQ at $\bar{\boldsymbol{x}}$.

Theorem

Let $\bar{\boldsymbol{x}}$ be a local minimum of MPCC and let MPCC-MFCQ hold at $\bar{\boldsymbol{x}}$. Then $\bar{\boldsymbol{x}}$ is C-stationary.

Unforunately, many feasible points with descent directions turn out to be C-stationary as well, so the criterion is considered a very weak one.

MPCC-Abadie CQ and MPCC-Guignard CQ

The **MPCC-linearized cone** is

$$L_{\text{MPCC}}(\mathcal{F}, \bar{\boldsymbol{x}}) = L(\mathcal{F}, \bar{\boldsymbol{x}}) \cap \left\{ \boldsymbol{d} \;\middle|\; (\nabla G_i(\bar{\boldsymbol{x}})^T \boldsymbol{d}) \cdot (\nabla H_i(\bar{\boldsymbol{x}})^T \boldsymbol{d}) = 0,\ i \in I_{00}(\bar{\boldsymbol{x}}) \right\}$$

and satisfies

$$T(\mathcal{F}, \bar{\boldsymbol{x}}) \subseteq L_{\text{MPCC}}(\mathcal{F}, \bar{\boldsymbol{x}}) \subseteq L(\mathcal{F}, \bar{\boldsymbol{x}}).$$

This motivates the definitions:
MPCC-ACQ holds at $\bar{\boldsymbol{x}} \in \mathcal{F}$ if

$$T(\mathcal{F}, \bar{\boldsymbol{x}}) = L_{\text{MPCC}}(\mathcal{F}, \bar{\boldsymbol{x}}).$$

MPCC-GCQ holds at $\bar{\boldsymbol{x}} \in \mathcal{F}$ if

$$T(\mathcal{F}, \bar{\boldsymbol{x}})^\circ = L_{\text{MPCC}}(\mathcal{F}, \bar{\boldsymbol{x}})^\circ.$$

Mordukhovich stationarity (M)

A point $\boldsymbol{x} \in \mathcal{F}$ is called **Mordukhovich-** or **M-stationarity** if there are multipliers $\boldsymbol{\lambda}_C$, $\boldsymbol{\mu}_D$, $\boldsymbol{\mu}_G$, $\boldsymbol{\mu}_H$ such that

$$\nabla \mathcal{L}_{\text{MPCC}}(\bar{\boldsymbol{x}}, \boldsymbol{\lambda}_C, \boldsymbol{\mu}_D, \boldsymbol{\mu}_G, \boldsymbol{\mu}_H) = \boldsymbol{0} \qquad \boldsymbol{\mu}_D \geqslant \boldsymbol{0}$$

$$\mu_{D,i} = 0 \text{ if } D_i(\bar{\boldsymbol{x}}) > 0$$

$$\mu_{G,i} \cdot \mu_{H,i} = 0 \text{ or } \mu_{G,i} \geqslant 0,\ \mu_{H,i} \geqslant 0 \text{ if } i \in I_{00}(\bar{\boldsymbol{x}})$$

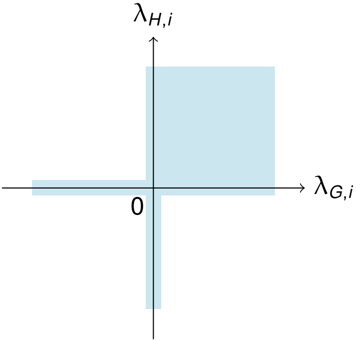
$$\mu_{H,i} = 0 \text{ if } i \in I_{0+}(\bar{\boldsymbol{x}})$$

$$\mu_{G,i} = 0 \text{ if } i \in I_{+0}(\bar{\boldsymbol{x}})$$

Theorem

If $\bar{\boldsymbol{x}}$ is a local minimum of MPCC, and MPCC-GCQ holds at $\bar{\boldsymbol{x}}$, then $\bar{\boldsymbol{x}}$ is an M-stationary point.

M-stationarity multiplier set



Optimal multiplier signs for M-stationarity.

Summary of Implication Chains

$$
\begin{tikzcd}
 & & \text{(always)} \arrow[r] & \text{B-stationarity} \\
\text{KKT} & \text{GCQ} \arrow[l] & \text{MPCC-LICQ} \arrow[l, Rightarrow] \arrow[r] \arrow[d, Rightarrow] & \text{S-stationarity} \arrow[u, Rightarrow] \arrow[d, Rightarrow] \\
 & \cancel{\text{ACQ}} \arrow[u, Rightarrow] & \text{MPCC-MFCQ} \arrow[d, Rightarrow] & \text{M-stationarity} \arrow[d, Rightarrow] \\
 & \cancel{\text{MFCQ}} \arrow[u, Rightarrow] & \text{MPCC-ACQ} \arrow[d, Rightarrow] & \text{C-stationarity} \\
 & \cancel{\text{LICQ}} \arrow[u, Rightarrow] & \text{MPCC-GCQ} \arrow[uur] &
\end{tikzcd}
$$

$\Longrightarrow$ “implies”

$\longrightarrow$ “is a necessary optimality condition”

I haven’t talked about A-, L-, T-, and W-stationarity ...