Levenberg-Marquardt Globalization of Newton-min for Complementarity Problems

Baptiste Plaquevent-Jourdain, with
Jean-Pierre Dussault, Université de Sherbrooke
Jean Charles Gilbert, INRIA Paris

July, 25th, 2024

Outline

Outline

Complementarity Problems

General form [FP03]

$F, G : \mathbb{R}^n \mapsto \mathbb{R}^n$ smooth vector functions,

$$\text{find } x \in \mathbb{R}^n \text{ s.t. : } F(x) \geq 0, G(x) \geq 0, F(x)^\mathsf{T} G(x) = 0$$
$$\Leftrightarrow 0 \leq F(x) \perp G(x) \geq 0. \quad (1)$$

Other forms/expressions exist:

- F or G is the identity ($G(x) = x$), $0 \leq F(x) \perp x \geq 0$.
- Linear Complementarity Problem [CPS92]

$$0 \leq x \perp (Mx + q) \geq 0. \quad (2)$$

Complementarity Problems

General form [FP03]

$F, G : \mathbb{R}^n \mapsto \mathbb{R}^n$ smooth vector functions,

$$\text{find } x \in \mathbb{R}^n \text{ s.t. : } F(x) \geq 0, G(x) \geq 0, F(x)^{\mathsf{T}} G(x) = 0 \\ \Leftrightarrow 0 \leq F(x) \perp G(x) \geq 0. \tag{1}$$

Other forms/expressions exist:

- F or G is the identity ($G(x) = x$), $0 \leq F(x) \perp x \geq 0$.
- Linear Complementarity Problem [CPS92]

$$0 \leq x \perp (Mx + q) \geq 0. \tag{2}$$

Reformulation trick: C-functions

The min C-function

One can reformulate (1) as finding $x \in \mathbb{R}^n$ such that

$$H_{\min}(x) := \min(F(x), G(x)) = (\min(F_i(x), G_i(x)))_{i \in [1:n]} = 0. \quad (3)$$

C-function: (1) $\Leftrightarrow H(x) = 0$, system of nonsmooth equations.

Another C-function is Fischer-Burmeister

$$\begin{aligned} \varphi_{\mathrm{FB}}(F_i(x), G_i(x)) &:= \sqrt{F_i(x)^2 + G_i(x)^2} - (F_i(x) + G_i(x)) \\ H_{\mathrm{FB}}(F(x), G(x)) &= (\varphi_{\mathrm{FB}}(F_i(x), G_i(x)))_{i \in [1:n]} = 0 \end{aligned} \quad (4)$$

(Many variants of φ_{FB}, parameter-based families...)

Reformulation trick: C-functions

The min C-function

One can reformulate (1) as finding $x \in \mathbb{R}^n$ such that

$$H_{\min}(x) := \min(F(x), G(x)) = (\min(F_i(x), G_i(x)))_{i\in[1:n]} = 0. \quad (3)$$

C-function: (1) $\Leftrightarrow H(x) = 0$, system of nonsmooth equations.

Another C-function is Fischer-Burmeister

$$\begin{aligned} \varphi_{\mathrm{FB}}(F_i(x), G_i(x)) &:= \sqrt{F_i(x)^2 + G_i(x)^2} - (F_i(x) + G_i(x)) \\ H_{\mathrm{FB}}(F(x), G(x)) &= (\varphi_{\mathrm{FB}}(F_i(x), G_i(x)))_{i\in[1:n]} = 0 \end{aligned} \quad (4)$$

(Many variants of φ_{FB}, parameter-based families...)

Comparison - 1

Minimum vs FB

	$H_{\min}(F,G)$	$H_{\mathrm{FB}}(F,G)$
formula(F,G)	piecewise linear✓	nonlinear
differentiable?	not if $F_i(x) = G_i(x)$, $F_i'(x) \neq G_i'(x)$	everywhere outside $F_i(x) = 0 = G_i(x)$ ✓

Semismooth Newton Method (SNM)

Adaptation of Newton's method to obtain fast local convergence.

General framework ($H_{\min}, H_{\mathrm{FB}}$)

- $k = 0, x^0 \in \mathbb{R}^n$, loop over k
- if $H(x^k) = 0$, stop
- solve $H(x^k) + A(x^k, d^k) = 0$; $A(x^k, \cdot)$ replaces "$H'(x^k)$"
- $x^{k+1} = x^k + d^k$

For instance, $A(x^k, d^k) = Jd^k$ for some $J \in \partial H(x^k)$.
With regularity assumptions to ensure each Jacobian is nonsingular.

Semismooth Newton Method (SNM)

Adaptation of Newton's method to obtain fast local convergence.

General framework ($H_{\min}$, H_{FB})

- $k = 0, x^0 \in \mathbb{R}^n$, loop over k
- if $H(x^k) = 0$, stop
- solve $H(x^k) + A(x^k, d^k) = 0$; $A(x^k, \cdot)$ replaces "$H'(x^k)$"
- $x^{k+1} = x^k + d^k$

For instance, $A(x^k, d^k) = Jd^k$ for some $J \in \partial H(x^k)$.
With regularity assumptions to ensure each Jacobian is nonsingular.

Comparison - 2

Minimum vs FB

	$H_{\min}(F,G)$	$H_{\mathrm{FB}}(F,G)$
formula(F,G)	piecewise linear ✓	nonlinear
differentiable?	not if $F_i(x) = G_i(x)$, $F_i'(x) \neq G_i'(x)$	everywhere outside $F_i(x) = 0 = G_i(x)$ ✓
SNM local convergence	if $\forall\ J \in \partial H_{\min}(x^*)$, J nonsingular: quadratic	if $\forall\ J \in \partial H_{\mathrm{FB}}(x^*)[*]$, J nonsingular: quadratic
bonus	finite if F, G affine	much more studied

$[*]$ But actually, more complicated to verify for $\partial H_{\mathrm{FB}}(x^*)$.
φ_{FB} "concentrates" the nondifferentiability in $(0,0)$.

Towards globalization

What about finding a suitable first iterate? $\rightarrow$ globalization
Use of a merit function $\theta = \frac{1}{2}H^{\mathsf{T}}H = ||H||^2/2$.

Focus of the talk:

convergence and globalization properties using H_{min}

Inspired from a polyhedral approach [DFG19] with linesearch.

Outline

1. Problem setting
2. Polyhedral approach
3. Bypassing regularity
   - Least-squares and regularization
   - Technical choice of the weights
4. Convergence
5. Appendices

The basic Newton-min algorithm

$$\text{smooth} \begin{cases} \mathcal{F}(x) := \{i : F_i(x) < G_i(x)\}, \\ \mathcal{G}(x) := \{i : F_i(x) > G_i(x)\}, \end{cases}$$

$$\text{(maybe) nonsmooth} \begin{cases} \mathcal{E}(x) := \{i : F_i(x) = G_i(x)\}. \end{cases}$$

The equality indices $\mathcal{E}(x)$ are partitioned into $\mathcal{E}(x) = \mathcal{E}_{\mathcal{F}}(x) \cup \mathcal{E}_{\mathcal{G}}(x)$, then

$$d(x) \text{ solution of } \begin{cases} (F(x) + F'(x)d)_{\mathcal{F}(x)\cup\mathcal{E}_{\mathcal{F}}(x)} & = 0 \\ (G(x) + G'(x)d)_{\mathcal{G}(x)\cup\mathcal{E}_{\mathcal{G}}(x)} & = 0 \end{cases}$$

- quite simple and can be very efficient;
- some theoretical convergence difficulties;
- which partition is essential; wrong choice can increase θ.

The basic Newton-min algorithm

$$\text{smooth} \begin{cases} \mathcal{F}(x) := \{i : F_i(x) < G_i(x)\}, \\ \mathcal{G}(x) := \{i : F_i(x) > G_i(x)\}, \end{cases}$$

$$\text{(maybe) nonsmooth} \begin{cases} \mathcal{E}(x) := \{i : F_i(x) = G_i(x)\}. \end{cases}$$

The equality indices $\mathcal{E}(x)$ are partitioned into $\mathcal{E}(x) = \mathcal{E}_{\mathcal{F}}(x) \cup \mathcal{E}_{\mathcal{G}}(x)$, then

$$d(x) \text{ solution of } \begin{cases} (F(x) + F'(x)d)_{\mathcal{F}(x)\cup\mathcal{E}_{\mathcal{F}}(x)} & = 0 \\ (G(x) + G'(x)d)_{\mathcal{G}(x)\cup\mathcal{E}_{\mathcal{G}}(x)} & = 0 \end{cases}$$

- quite simple and can be very efficient;
- some theoretical convergence difficulties;
- which partition is essential; wrong choice can increase θ.

Local treatment

$(F + F'd)_{\mathcal{F}(x)} = 0, (G + G'd)_{\mathcal{G}(x)} = 0$, '$(x)$' dropped

$$\theta'(x; d) := \sum_{i \in \mathcal{F}(x)} F_i F_i' d + \sum_{i \in \mathcal{G}(x)} G_i G_i' d + \sum_{i \in \mathcal{E}(x)} H_i \min(F_i' d, G_i' d)$$

$$= \underbrace{-2\theta(x)}_{\text{smooth}} + \sum_{i \in \mathcal{E}(x)} \underbrace{H_i(\min(F_i + F_i' d, G_i + G_i' d))}_{\text{nonsmooth}}$$

Let $\mathcal{E}^{0+}(x) := \{i : F_i = G_i \geq 0\}$ and $\mathcal{E}^{-}(x) := \{i : F_i = G_i < 0\}$, $\mathcal{E}^{0+}_{\mathcal{F}}(x) \cup \mathcal{E}^{0+}_{\mathcal{G}}(x)$ be a partition of $\mathcal{E}^{0+}(x)$: polyhedron in d

$$\begin{cases} F_i + F_i' d = 0 & i \in \mathcal{F}(x) \cup \mathcal{E}^{0+}_{\mathcal{F}}(x) \\ G_i + G_i' d = 0 & i \in \mathcal{G}(x) \cup \mathcal{E}^{0+}_{\mathcal{G}}(x) \\ F_i + F_i' d \geq 0 & i \in \mathcal{E}^{-}(x) \\ G_i + G_i' d \geq 0 & i \in \mathcal{E}^{-}(x) \end{cases} \Rightarrow \theta'(x; d) \leq -2\theta(x)$$

Local treatment

$(F+F'd)_{\mathcal{F}(x)}=0, (G+G'd)_{\mathcal{G}(x)}=0$, '$(x)$' dropped

$$\theta'(x;d):=\sum_{i\in\mathcal{F}(x)} F_iF_i'd+\sum_{i\in\mathcal{G}(x)} G_iG_i'd+\sum_{i\in\mathcal{E}(x)} H_i\min(F_i'd,G_i'd)$$
$$=\underbrace{-2\theta(x)}_{\text{smooth}}+\sum_{i\in\mathcal{E}(x)}\underbrace{H_i(\min(F_i+F_i'd,G_i+G_i'd))}_{\text{nonsmooth}}$$

Let $\mathcal{E}^{0+}(x):=\{i : F_i=G_i\geq 0\}$ and $\mathcal{E}^{-}(x):=\{i : F_i=G_i<0\}$, $\mathcal{E}^{0+}_{\mathcal{F}}(x)\cup\mathcal{E}^{0+}_{\mathcal{G}}(x)$ be a partition of $\mathcal{E}^{0+}(x)$: polyhedron in d

$$\left\{\begin{array}{ll} F_i+F_i'd=0 & i\in\mathcal{F}(x)\cup\mathcal{E}^{0+}_{\mathcal{F}}(x)\\ G_i+G_i'd=0 & i\in\mathcal{G}(x)\cup\mathcal{E}^{0+}_{\mathcal{G}}(x)\\ F_i+F_i'd\geq 0 & i\in\mathcal{E}^{-}(x)\\ G_i+G_i'd\geq 0 & i\in\mathcal{E}^{-}(x)\end{array}\right. \Rightarrow \theta'(x;d)\leq -2\theta(x)$$

From local to global

Key steps of the polyhedral linesearch version

- near a negative kink, $F_i(x) \lesssim G_i(x)$, F_i used
- but in $x + d$, it may be $G_i(x+d) \lesssim F_i(x+d)$
- $F_i(x) = G_i(x) \to |F_i(x) - G_i(x)| \leq \tau$ for $\tau > 0$ small
- regularity assumptions for $\exists\ d$ bounded
- global convergence obtained with linesearch

Outline

Outline

1. Problem setting
2. Polyhedral approach
3. Bypassing regularity
   - Least-squares and regularization
   - Technical choice of the weights
4. Convergence
5. Appendices

What if the polyhedron is empty?

If suitable regularity, polyhedrons are non-empty so there is a d.
Otherwise, least-squares to find "a best possible d".

Recall the polyhedral system

$$\begin{cases} F_i + F_i' d = 0 & i \in \mathcal{F}(x) \cup \mathcal{E}_{\mathcal{F}}^{0+}(x) \\ G_i + G_i' d = 0 & i \in \mathcal{G}(x) \cup \mathcal{E}_{\mathcal{G}}^{0+}(x) \\ F_i + F_i' d \geq 0 & i \in \mathcal{E}^{-}(x) \\ G_i + G_i' d \geq 0 & i \in \mathcal{E}^{-}(x) \end{cases} \tag{5}$$

Formalism with weights

The $\mathcal{E}_{\mathcal{F}}^{0+}(x) \cup \mathcal{E}_{\mathcal{G}}^{0+}(x)$ part, for $\gamma_{\mathcal{E}^{0+}(x)} \in \{0,1\}^{\mathcal{E}^{0+}(x)}$, $\overline{\gamma}_i := 1 - \gamma_i$

$$\gamma_i(F_i(x) + F_i'(x)d) + \overline{\gamma}_i(G_i(x) + G_i'(x)d) = 0$$

What if the polyhedron is empty?

If suitable regularity, polyhedrons are non-empty so there is a d.
Otherwise, least-squares to find "a best possible d".

Recall the polyhedral system

$$\begin{cases} F_i + F_i' d = 0 & i \in \mathcal{F}(x) \cup \mathcal{E}_{\mathcal{F}}^{0+}(x) \\ G_i + G_i' d = 0 & i \in \mathcal{G}(x) \cup \mathcal{E}_{\mathcal{G}}^{0+}(x) \\ F_i + F_i' d \geq 0 & i \in \mathcal{E}^{-}(x) \\ G_i + G_i' d \geq 0 & i \in \mathcal{E}^{-}(x) \end{cases} \tag{5}$$

Formalism with weights

The $\mathcal{E}_{\mathcal{F}}^{0+}(x) \cup \mathcal{E}_{\mathcal{G}}^{0+}(x)$ part, for $\gamma_{\mathcal{E}^{0+}(x)} \in \{0,1\}^{\mathcal{E}^{0+}(x)}$, $\overline{\gamma}_i := 1 - \gamma_i$

$$\gamma_i (F_i(x) + F_i'(x) d) + \overline{\gamma}_i (G_i(x) + G_i'(x) d) = 0$$

Least-squares formulation

$$\min_{d\in\mathbb{R}^n} q(x,d)/2,$$

$$q(x,d) = \sum_{i\in\mathcal{F}(x)} (F_i(x)+F_i'(x)d)^2 + \sum_{i\in\mathcal{G}(x)} (G_i(x)+G_i'(x)d)^2$$
$$+ \sum_{i\in\mathcal{E}^{0+}(x)} \gamma_i(F_i(x)+F_i'(x)d)^2 + \overline{\gamma}_i(G_i(x)+G_i'(x)d)^2 \qquad (6)$$
$$+ \sum_{i\in\mathcal{E}^{-}(x)} \gamma_i[(F_i(x)+F_i'(x)d)^-]^2 + \overline{\gamma}_i[(G_i(x)+G_i'(x)d)^-]^2$$

Twice the $i\in\mathcal{E}^-(x)$: $\gamma_{\mathcal{E}^-(x)} \in [0,1]^{\mathcal{E}^-(x)}$ (see later).
Levenberg-Marquardt (LM) regularization - convex diff (not C^1)

$$\min_{d\in\mathbb{R}^n} \varphi_x(d) := \frac{1}{2}[q(x,d)+\lambda d^\mathsf{T} S d], \quad \lambda\geq 0, S\succ 0 \qquad (7)$$

Least-squares formulation

$$\min_{d \in \mathbb{R}^n} q(x, d)/2,$$

$$q(x,d) = \sum_{i \in \mathcal{F}(x)} (F_i(x) + F_i'(x)d)^2 + \sum_{i \in \mathcal{G}(x)} (G_i(x) + G_i'(x)d)^2$$
$$+ \sum_{i \in \mathcal{E}^{0+}(x)} \gamma_i (F_i(x) + F_i'(x)d)^2 + \overline{\gamma}_i (G_i(x) + G_i'(x)d)^2 \qquad (6)$$
$$+ \sum_{i \in \mathcal{E}^{-}(x)} \gamma_i [(F_i(x) + F_i'(x)d)^-]^2 + \overline{\gamma}_i [(G_i(x) + G_i'(x)d)^-]^2$$

Twice the $i \in \mathcal{E}^-(x)$: $\gamma_{\mathcal{E}^-(x)} \in [0,1]^{\mathcal{E}^-(x)}$ (see later).
Levenberg-Marquardt (LM) regularization - convex diff (not C^1)

$$\min_{d \in \mathbb{R}^n} \varphi_x(d) := \frac{1}{2}[q(x,d) + \lambda d^\top S d], \quad \lambda \geq 0, S \succ 0 \qquad (7)$$

Least-squares formulation

$$\min_{d\in\mathbb{R}^n} q(x,d)/2,$$

$$q(x,d) = \sum_{i\in\mathcal{F}(x)} (F_i(x)+F_i'(x)d)^2 + \sum_{i\in\mathcal{G}(x)} (G_i(x)+G_i'(x)d)^2$$
$$+ \sum_{i\in\mathcal{E}^{0+}(x)} \gamma_i(F_i(x)+F_i'(x)d)^2 + \overline{\gamma}_i(G_i(x)+G_i'(x)d)^2 \qquad (6)$$
$$+ \sum_{i\in\mathcal{E}^{-}(x)} \gamma_i[(F_i(x)+F_i'(x)d)^-]^2 + \overline{\gamma}_i[(G_i(x)+G_i'(x)d)^-]^2$$

Twice the $i\in\mathcal{E}^-(x)$: $\gamma_{\mathcal{E}^-(x)} \in [0,1]^{\mathcal{E}^-(x)}$ (see later).
Levenberg-Marquardt (LM) regularization - convex diff (not C^1)

$$\min_{d\in\mathbb{R}^n} \varphi_x(d) := \frac{1}{2}[q(x,d)+\lambda d^\mathsf{T} S d], \quad \lambda\geq 0, S\succ 0 \qquad (7)$$

Least-squares formulation

$$\min_{d\in\mathbb{R}^n} q(x,d)/2,$$

$$q(x,d) = \sum_{i\in\mathcal{F}(x)} (F_i(x)+F_i'(x)d)^2 + \sum_{i\in\mathcal{G}(x)} (G_i(x)+G_i'(x)d)^2$$
$$+ \sum_{i\in\mathcal{E}^{0+}(x)} \gamma_i(F_i(x)+F_i'(x)d)^2 + \overline{\gamma}_i(G_i(x)+G_i'(x)d)^2 \tag{6}$$
$$+ \sum_{i\in\mathcal{E}^{-}(x)} \gamma_i[(F_i(x)+F_i'(x)d)^-]^2 + \overline{\gamma}_i[(G_i(x)+G_i'(x)d)^-]^2$$

Twice the $i\in\mathcal{E}^-(x)$: $\gamma_{\mathcal{E}^-(x)} \in [0,1]^{\mathcal{E}^-(x)}$ (see later).
Levenberg-Marquardt (LM) regularization - convex diff (not C^1)

$$\min_{d\in\mathbb{R}^n} \varphi_x(d) := \frac{1}{2}[q(x,d)+\lambda d^\mathsf{T} S d], \quad \lambda\geq 0, S\succ 0 \tag{7}$$

Least-squares formulation - goal

main elements

- φ_x serves as a piecewise quadratic model of θ at x
- φ_x always has a minimizer d even if the polyhedron is empty
- convex relaxation: F_i and G_i for $i \in \mathcal{E}^-(x) \cup \mathcal{E}^{0+}(x)$
- $g(\gamma_{\mathcal{E}^{0+}(x)}, \gamma_{\mathcal{E}^-(x)}) := \nabla \varphi_x(d=0)$ has a descent property?
- is there a way to characterize stationarity of θ via φ_x?

Let $\Gamma_{\mathcal{E}^{0+}(x)} = \mathrm{Diag}(\gamma_{\mathcal{E}^{0+}(x)}), \Gamma_{\mathcal{E}^-(x)} = \mathrm{Diag}(\gamma_{\mathcal{E}^-(x)})$

$$\begin{aligned} g(\gamma) = & \nabla F_{\mathcal{F}(x)}(x)^T F_{\mathcal{F}(x)}(x) + \nabla G_{\mathcal{G}(x)}(x)^T G_{\mathcal{G}(x)}(x) \\ & + [\nabla F_{\mathcal{E}^{0+}(x)}(x)^T \Gamma_{\mathcal{E}^{0+}(x)} + \nabla G_{\mathcal{E}^{0+}(x)}(x)^T \overline{\Gamma}_{\mathcal{E}^{0+}(x)}] H_{\mathcal{E}^{0+}(x)}(x) \\ & + [\nabla F_{\mathcal{E}^-(x)}(x)^T \Gamma_{\mathcal{E}^-(x)} + \nabla G_{\mathcal{E}^-(x)}(x)^T \overline{\Gamma}_{\mathcal{E}^-(x)}] H_{\mathcal{E}^-(x)}(x) \\ & (F_{\mathcal{E}(x)}(x) = H_{\mathcal{E}(x)}(x) = G_{\mathcal{E}(x)}(x)) \end{aligned}$$

Least-squares formulation - goal

main elements

- φ_x serves as a piecewise quadratic model of θ at x
- φ_x always has a minimizer d even if the polyhedron is empty
- convex relaxation: F_i and G_i for $i \in \mathcal{E}^-(x) \cup \mathcal{E}^{0+}(x)$
- $g(\gamma_{\mathcal{E}^{0+}(x)}, \gamma_{\mathcal{E}^-(x)}) := \nabla\varphi_x(d=0)$ has a descent property?
- is there a way to characterize stationarity of θ via φ_x?

Let $\Gamma_{\mathcal{E}^{0+}(x)} = \mathrm{Diag}(\gamma_{\mathcal{E}^{0+}(x)}), \Gamma_{\mathcal{E}^-(x)} = \mathrm{Diag}(\gamma_{\mathcal{E}^-(x)})$

$$
\begin{aligned}
g(\gamma) = & \nabla F_{\mathcal{F}(x)}(x)^\top F_{\mathcal{F}(x)}(x) + \nabla G_{\mathcal{G}(x)}(x)^\top G_{\mathcal{G}(x)}(x) \\
& + [\nabla F_{\mathcal{E}^{0+}(x)}(x)^\top \Gamma_{\mathcal{E}^{0+}(x)} + \nabla G_{\mathcal{E}^{0+}(x)}(x)^\top \overline{\Gamma}_{\mathcal{E}^{0+}(x)}] H_{\mathcal{E}^{0+}(x)}(x) \\
& + [\nabla F_{\mathcal{E}^-(x)}(x)^\top \Gamma_{\mathcal{E}^-(x)} + \nabla G_{\mathcal{E}^-(x)}(x)^\top \overline{\Gamma}_{\mathcal{E}^-(x)}] H_{\mathcal{E}^-(x)}(x) \\
& (F_{\mathcal{E}(x)}(x) = H_{\mathcal{E}(x)}(x) = G_{\mathcal{E}(x)}(x))
\end{aligned}
$$

Least-squares formulation - goal

main elements

- φ_x serves as a piecewise quadratic model of θ at x
- φ_x always has a minimizer d even if the polyhedron is empty
- convex relaxation: F_i and G_i for $i \in \mathcal{E}^-(x) \cup \mathcal{E}^{0+}(x)$
- $g(\gamma_{\mathcal{E}^{0+}(x)}, \gamma_{\mathcal{E}^-(x)}) := \nabla\varphi_x(d=0)$ has a descent property?
- is there a way to characterize stationarity of θ via φ_x?

Let $\Gamma_{\mathcal{E}^{0+}(x)} = \mathrm{Diag}(\gamma_{\mathcal{E}^{0+}(x)}), \Gamma_{\mathcal{E}^-(x)} = \mathrm{Diag}(\gamma_{\mathcal{E}^-(x)})$

$$
\begin{aligned}
g(\gamma) = &\nabla F_{\mathcal{F}(x)}(x)^{\mathsf{T}} F_{\mathcal{F}(x)}(x) + \nabla G_{\mathcal{G}(x)}(x)^{\mathsf{T}} G_{\mathcal{G}(x)}(x) \\
&+ [\nabla F_{\mathcal{E}^{0+}(x)}(x)^{\mathsf{T}} \Gamma_{\mathcal{E}^{0+}(x)} + \nabla G_{\mathcal{E}^{0+}(x)}(x)^{\mathsf{T}} \overline{\Gamma}_{\mathcal{E}^{0+}(x)}] H_{\mathcal{E}^{0+}(x)}(x) \\
&+ [\nabla F_{\mathcal{E}^-(x)}(x)^{\mathsf{T}} \Gamma_{\mathcal{E}^-(x)} + \nabla G_{\mathcal{E}^-(x)}(x)^{\mathsf{T}} \overline{\Gamma}_{\mathcal{E}^-(x)}] H_{\mathcal{E}^-(x)}(x) \\
&(F_{\mathcal{E}(x)}(x) = H_{\mathcal{E}(x)}(x) = G_{\mathcal{E}(x)}(x))
\end{aligned}
$$

Least-squares formulation - goal

main elements

- φ_x serves as a piecewise quadratic model of θ at x
- φ_x always has a minimizer d even if the polyhedron is empty
- convex relaxation: F_i and G_i for $i \in \mathcal{E}^-(x) \cup \mathcal{E}^{0+}(x)$
- $g(\gamma_{\mathcal{E}^{0+}(x)}, \gamma_{\mathcal{E}^-(x)}) := \nabla\varphi_x(d=0)$ has a descent property?
- is there a way to characterize stationarity of θ via φ_x?

Let $\Gamma_{\mathcal{E}^{0+}(x)} = \mathrm{Diag}(\gamma_{\mathcal{E}^{0+}(x)}), \Gamma_{\mathcal{E}^-(x)} = \mathrm{Diag}(\gamma_{\mathcal{E}^-(x)})$

$$\begin{aligned} g(\gamma) = & \nabla F_{\mathcal{F}(x)}(x)^\mathsf{T} F_{\mathcal{F}(x)}(x) + \nabla G_{\mathcal{G}(x)}(x)^\mathsf{T} G_{\mathcal{G}(x)}(x) \\ & + [\nabla F_{\mathcal{E}^{0+}(x)}(x)^\mathsf{T} \Gamma_{\mathcal{E}^{0+}(x)} + \nabla G_{\mathcal{E}^{0+}(x)}(x)^\mathsf{T} \overline{\Gamma}_{\mathcal{E}^{0+}(x)}] H_{\mathcal{E}^{0+}(x)}(x) \\ & + [\nabla F_{\mathcal{E}^-(x)}(x)^\mathsf{T} \Gamma_{\mathcal{E}^-(x)} + \nabla G_{\mathcal{E}^-(x)}(x)^\mathsf{T} \overline{\Gamma}_{\mathcal{E}^-(x)}] H_{\mathcal{E}^-(x)}(x) \\ & (F_{\mathcal{E}(x)}(x) = H_{\mathcal{E}(x)}(x) = G_{\mathcal{E}(x)}(x)) \end{aligned}$$

Least-squares formulation - goal

main elements

- φ_x serves as a piecewise quadratic model of θ at x
- φ_x always has a minimizer d even if the polyhedron is empty
- convex relaxation: F_i and G_i for $i \in \mathcal{E}^-(x) \cup \mathcal{E}^{0+}(x)$
- $g(\gamma_{\mathcal{E}^{0+}(x)}, \gamma_{\mathcal{E}^-(x)}) := \nabla\varphi_x(d = 0)$ has a descent property?
- is there a way to characterize stationarity of θ via φ_x?

Let $\Gamma_{\mathcal{E}^{0+}(x)} = \mathrm{Diag}(\gamma_{\mathcal{E}^{0+}(x)}), \Gamma_{\mathcal{E}^-(x)} = \mathrm{Diag}(\gamma_{\mathcal{E}^-(x)})$

$$\begin{aligned} g(\gamma) = &\nabla F_{\mathcal{F}(x)}(x)^\mathsf{T} F_{\mathcal{F}(x)}(x) + \nabla G_{\mathcal{G}(x)}(x)^\mathsf{T} G_{\mathcal{G}(x)}(x) \\ &+ [\nabla F_{\mathcal{E}^{0+}(x)}(x)^\mathsf{T}\Gamma_{\mathcal{E}^{0+}(x)} + \nabla G_{\mathcal{E}^{0+}(x)}(x)^\mathsf{T}\overline{\Gamma}_{\mathcal{E}^{0+}(x)}] H_{\mathcal{E}^{0+}(x)}(x) \\ &+ [\nabla F_{\mathcal{E}^-(x)}(x)^\mathsf{T}\Gamma_{\mathcal{E}^-(x)} + \nabla G_{\mathcal{E}^-(x)}(x)^\mathsf{T}\overline{\Gamma}_{\mathcal{E}^-(x)}] H_{\mathcal{E}^-(x)}(x) \\ &(F_{\mathcal{E}(x)}(x) = H_{\mathcal{E}(x)}(x) = G_{\mathcal{E}(x)}(x)) \end{aligned}$$

Outline

1. Problem setting
2. Polyhedral approach
3. Bypassing regularity
   - Least-squares and regularization
   - Technical choice of the weights
4. Convergence
5. Appendices

Computations

One wants $\theta'(x; -g(\gamma_{\mathcal{E}^{0+}(x)}, \gamma_{\mathcal{E}^{-}(x)})) \leq 0$.
Let $\gamma_+ := \gamma_{\mathcal{E}^{0+}(x)}, \gamma_- := \gamma_{\mathcal{E}^{-}(x)}$, $\Gamma_+ = \mathrm{Diag}(\gamma_+), \Gamma_- = \mathrm{Diag}(\gamma_-)$

$$\theta'(x; -g(\gamma_+, \gamma_-)) = -||g(\gamma_+, \gamma_-)||^2$$

$$0 \geq \; + H_{\mathcal{E}^{0+}(x)}(x)^\top \underbrace{[\min(-F'_{\mathcal{E}^{0+}(x)}(x)g(\gamma_+, \gamma_-), -G'_{\mathcal{E}^{0+}(x)}(x)g(\gamma_+, \gamma_-)) + \Gamma_+ F'_{\mathcal{E}^{0+}(x)}(x)g(\gamma_+, \gamma_-) + \overline{\Gamma}_+ G'_{\mathcal{E}^{0+}(x)}(x)g(\gamma_+, \gamma_-)]}_{\leq 0}$$

$$0 \leq \; + H_{\mathcal{E}^{-}(x)}(x)^\top \underbrace{[\min(-F'_{\mathcal{E}^{-}(x)}(x)g(\gamma_+, \gamma_-), -G'_{\mathcal{E}^{-}(x)}(x)g(\gamma_+, \gamma_-)) + \Gamma_- F'_{\mathcal{E}^{-}(x)}(x)g(\gamma_+, \gamma_-) + \overline{\Gamma}_- G'_{\mathcal{E}^{-}(x)}(x)g(\gamma_+, \gamma_-)]}_{\leq 0}$$

$$\min(-a, -b) + \gamma a + \overline{\gamma} b = \begin{cases} \overline{\gamma}(b-a) \leq 0 & \text{if } a \geq b \\ \gamma(a-b) \leq 0 & \text{if } a \leq b \end{cases}.$$

Computations

One wants $\theta'(x; -g(\gamma_{\mathcal{E}^{0+}(x)}, \gamma_{\mathcal{E}^-(x)})) \leq 0$.

Let $\gamma_+ := \gamma_{\mathcal{E}^{0+}(x)}, \gamma_- := \gamma_{\mathcal{E}^-(x)}$, $\Gamma_+ = \mathrm{Diag}(\gamma_+), \Gamma_- = \mathrm{Diag}(\gamma_-)$

$$\theta'(x; -g(\gamma_+, \gamma_-)) = -||g(\gamma_+, \gamma_-)||^2$$

$$0 \geq + H_{\mathcal{E}^{0+}(x)}(x)^{\mathsf{T}} \underbrace{[\min(-F'_{\mathcal{E}^{0+}(x)}(x) g(\gamma_+, \gamma_-), -G'_{\mathcal{E}^{0+}(x)}(x) g(\gamma_+, \gamma_-)) + \Gamma_+ F'_{\mathcal{E}^{0+}(x)}(x) g(\gamma_+, \gamma_-) + \overline{\Gamma}_+ G'_{\mathcal{E}^{0+}(x)}(x) g(\gamma_+, \gamma_-)]}_{\leq 0}$$

$$0 \leq + H_{\mathcal{E}^-(x)}(x)^{\mathsf{T}} \underbrace{[\min(-F'_{\mathcal{E}^-(x)}(x) g(\gamma_+, \gamma_-), -G'_{\mathcal{E}^-(x)}(x) g(\gamma_+, \gamma_-)) + \Gamma_- F'_{\mathcal{E}^-(x)}(x) g(\gamma_+, \gamma_-) + \overline{\Gamma}_- G'_{\mathcal{E}^-(x)}(x) g(\gamma_+, \gamma_-)]}_{\leq 0}$$

$$\min(-a, -b) + \gamma a + \overline{\gamma} b = \begin{cases} \overline{\gamma}(b-a) \leq 0 & \text{if } a \geq b \\ \gamma(a-b) \leq 0 & \text{if } a \leq b \end{cases}.$$

Computations

One wants $\theta'(x; -g(\gamma_{\mathcal{E}^{0+}(x)}, \gamma_{\mathcal{E}^{-}(x)})) \leq 0$.

Let $\gamma_+ := \gamma_{\mathcal{E}^{0+}(x)}, \gamma_- := \gamma_{\mathcal{E}^{-}(x)}$, $\Gamma_+ = \mathrm{Diag}(\gamma_+), \Gamma_- = \mathrm{Diag}(\gamma_-)$

$$\theta'(x; -g(\gamma_+, \gamma_-)) = -||g(\gamma_+, \gamma_-)||^2$$

$$0 \geq + H_{\mathcal{E}^{0+}(x)}(x)^T \underbrace{[\min(-F'_{\mathcal{E}^{0+}(x)}(x)g(\gamma_+, \gamma_-), -G'_{\mathcal{E}^{0+}(x)}(x)g(\gamma_+, \gamma_-)) + \Gamma_+ F'_{\mathcal{E}^{0+}(x)}(x)g(\gamma_+, \gamma_-) + \overline{\Gamma}_+ G'_{\mathcal{E}^{0+}(x)}(x)g(\gamma_+, \gamma_-)]}_{\leq 0}$$

$$0 \leq + H_{\mathcal{E}^{-}(x)}(x)^T \underbrace{[\min(-F'_{\mathcal{E}^{-}(x)}(x)g(\gamma_+, \gamma_-), -G'_{\mathcal{E}^{-}(x)}(x)g(\gamma_+, \gamma_-)) + \Gamma_- F'_{\mathcal{E}^{-}(x)}(x)g(\gamma_+, \gamma_-) + \overline{\Gamma}_- G'_{\mathcal{E}^{-}(x)}(x)g(\gamma_+, \gamma_-)]}_{\leq 0}$$

$$\min(-a, -b) + \gamma a + \overline{\gamma} b = \begin{cases} \overline{\gamma}(b - a) \leq 0 & \text{if } a \geq b \\ \gamma(a - b) \leq 0 & \text{if } a \leq b \end{cases}.$$

Computations

One wants $\theta'(x; -g(\gamma_{\mathcal{E}^{0+}(x)}, \gamma_{\mathcal{E}^{-}(x)})) \leq 0$.

Let $\gamma_+ := \gamma_{\mathcal{E}^{0+}(x)}, \gamma_- := \gamma_{\mathcal{E}^{-}(x)}$, $\Gamma_+ = \mathrm{Diag}(\gamma_+), \Gamma_- = \mathrm{Diag}(\gamma_-)$

$$\theta'(x; -g(\gamma_+, \gamma_-)) = -||g(\gamma_+, \gamma_-)||^2$$

$$0 \geq \; + H_{\mathcal{E}^{0+}(x)}(x)^\mathsf{T}[\min(-F'_{\mathcal{E}^{0+}(x)}(x)g(\gamma_+, \gamma_-), -G'_{\mathcal{E}^{0+}(x)}(x)g(\gamma_+, \gamma_-)) \underbrace{+\Gamma_+ F'_{\mathcal{E}^{0+}(x)}(x)g(\gamma_+, \gamma_-) + \overline{\Gamma}_+ G'_{\mathcal{E}^{0+}(x)}(x)g(\gamma_+, \gamma_-)]}_{\leq 0}$$

$$0 \leq \; + H_{\mathcal{E}^{-}(x)}(x)^\mathsf{T}[\min(-F'_{\mathcal{E}^{-}(x)}(x)g(\gamma_+, \gamma_-), -G'_{\mathcal{E}^{-}(x)}(x)g(\gamma_+, \gamma_-)) \underbrace{+\Gamma_- F'_{\mathcal{E}^{-}(x)}(x)g(\gamma_+, \gamma_-) + \overline{\Gamma}_- G'_{\mathcal{E}^{-}(x)}(x)g(\gamma_+, \gamma_-)]}_{\leq 0}$$

$$\min(-a, -b) + \gamma a + \overline{\gamma} b = \begin{cases} \overline{\gamma}(b-a) \leq 0 & \text{if } a \geq b \\ \gamma(a-b) \leq 0 & \text{if } a \leq b \end{cases}.$$

Comments on the weights

Observations:

- the term in $\mathcal{E}^{0+}(x)$: always ≤ 0;
- the term in $\mathcal{E}^{-}(x)$: always ≥ 0;
- but both intertwine since $g(\gamma_+, \gamma_-)$

We present a way to have the ≥ 0 be 0:

$$\theta'(x; -g(\gamma_+, \gamma_-)) = -||g(\gamma_+, \gamma_-)||^2 - \cdots + 0 \leq -||g||^2 \leq 0$$

The wrong choice of γ_- can lead to $\theta'(x; -g(\gamma_+, \gamma_-)) \geq 0$!

Choosing correct weights

Lemma ((partial) Choice of the weights)

Let $\gamma_+ \in [0,1]^{\mathcal{E}^{0+}(x)}$ *be fixed. There exists a* $\gamma_-(\gamma_+)$ *such that the* ≥ 0 *term is* $= 0$*, i.e.,* $g = g(\gamma_+, \gamma_-(\gamma_+))$ *verifies*

$$
\begin{aligned}
&-\Gamma_- F'_{\mathcal{E}^-(x)}(x) g(\gamma_+, \gamma_-(\gamma_+)) - \overline{\Gamma}_- G'_{\mathcal{E}^-(x)}(x) g(\gamma_+, \gamma_-(\gamma_+)) \\
&= \min(-F'_{\mathcal{E}^-(x)}(x) g(\gamma_+, \gamma_-(\gamma_+)), -G'_{\mathcal{E}^-(x)}(x) g(\gamma_+, \gamma_-(\gamma_+)))
\end{aligned}
$$

- Quadratic equation since $g(\gamma_+, \gamma_-)$ is affine in γ
- $\gamma_+ \rightrightarrows \gamma_-(\gamma_+)$ may be multi-valued, $g(\gamma_+, \gamma_-(\gamma_+))$ unchanged
- $\mathcal{E}^{0+}(x) = \varnothing$: no "control" from γ_+
- if $g(\gamma_+, \gamma_-) = 0$ for some γ_-, trivial but not useful

Choosing correct weights

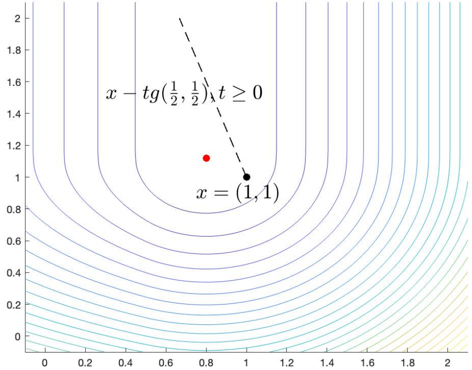
Lemma ((partial) Choice of the weights)

Let $\gamma_+ \in [0,1]^{\mathcal{E}^{0+}(x)}$ be fixed. There exists a $\gamma_-(\gamma_+)$ such that the ≥ 0 term is $= 0$, i.e., $g = g(\gamma_+, \gamma_-(\gamma_+))$ verifies

$$
\begin{aligned}
&-\Gamma_- F'_{\mathcal{E}^-(x)}(x) g(\gamma_+, \gamma_-(\gamma_+)) - \overline{\Gamma}_- G'_{\mathcal{E}^-(x)}(x) g(\gamma_+, \gamma_-(\gamma_+)) \\
&= \min(-F'_{\mathcal{E}^-(x)}(x) g(\gamma_+, \gamma_-(\gamma_+)), -G'_{\mathcal{E}^-(x)}(x) g(\gamma_+, \gamma_-(\gamma_+)))
\end{aligned}
$$

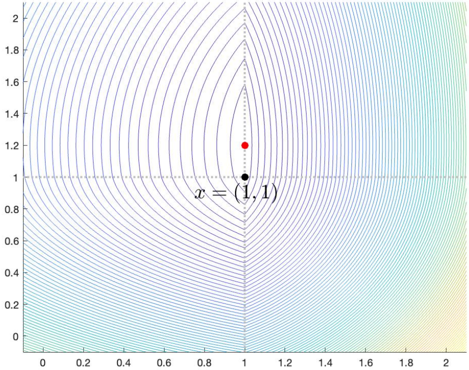
- Quadratic equation since $g(\gamma_+, \gamma_-)$ is affine in γ
- $\gamma_+ \rightrightarrows \gamma_-(\gamma_+)$ may be multi-valued, $g(\gamma_+, \gamma_-(\gamma_+))$ unchanged
- $\mathcal{E}^{0+}(x) = \varnothing$: no "control" from γ_+
- if $g(\gamma_+, \gamma_-) = 0$ for some γ_-, trivial but not useful

Illustration of the lemma - 1



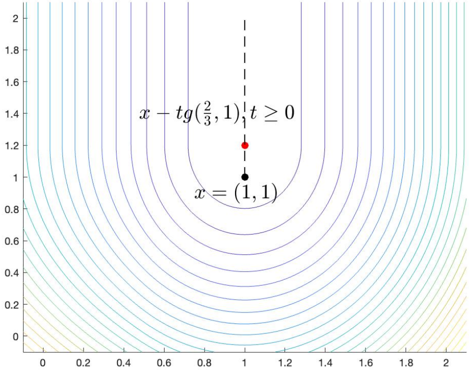
Level sets of φ_x with arbitrary weights $\gamma = (1/2, 1/2)$.

Illustration of the lemma - 2



Level sets of θ. The grey dotted lines are the kinks.

Illustration of the lemma - 3



Level sets of φ_x with the weights given by the lemma.

Summary

- Generalization of the polyhedral system;
- $\mathcal{E}^{0+}(x)$ and $\mathcal{E}^{-}(x)$ have different roles;
- convex combination(F_i, G_i) for the non-differentiable part.

Remaining questions

- ensure a descent property ($g(\gamma_+, \gamma_-) \neq 0$)
- stationarity of an iterate
- convergence

Summary

- Generalization of the polyhedral system;
- $\mathcal{E}^{0+}(x)$ and $\mathcal{E}^{-}(x)$ have different roles;
- convex combination(F_i, G_i) for the non-differentiable part.

Remaining questions

- ensure a descent property ($g(\gamma_+, \gamma_-) \neq 0$)
- stationarity of an iterate
- convergence

Outline

Stationarity detection with the weights

Proposition (characterizing stationarity)

The following properties are equivalent:

1) x *is* θ*-stationary,*
2) *for any* $\gamma_+ \in [0,1]^{\mathcal{E}^{0+}(x)}$, *let a* $\gamma_-(\gamma_+)$ *be given by the lemma, one has* $g(\gamma_+, \gamma_-(\gamma_+)) = 0$.

(Lineasearch: partition of $\mathcal{E}^{0+}(x) \Leftrightarrow \gamma_+ \in \{0,1\}^{\mathcal{E}^{0+}(x)}$.)

An iteration of the algorithm

Algorithm Substep for γ then usual LM substeps

```
Test: x^k is stationary or obtain a suitable γ_+
Weights computation: obtain γ_−(γ_+)
Get d(λ) = arg min_d φ_{x^k}(d) using λ, γ_+, γ_−, S_{k+1}
while Descent condition not satisfied do
    increase λ and recompute d(λ)
end while
Update x^{k+1} = x^k + d(λ) (and S_{k+1})
if Stronger descent condition is satisfied then
    decrease λ
end if
```

Main costs: γ_+, $\gamma_-(\gamma_+)$ (once per k), especially $d(\lambda)$.

Convergence properties

Proposition (sufficient decrease)

1) $d_k(\lambda)/||d_k(\lambda)|| \underset{\lambda \to +\infty}{\to} -S_k^{-1} g_k/||S_k^{-1} g||$
2) *for λ large enough, the descent condition holds*

Theorem (Convergence)

Let (x_k, λ_k, S_k) be a sequence generated by algorithm 1.

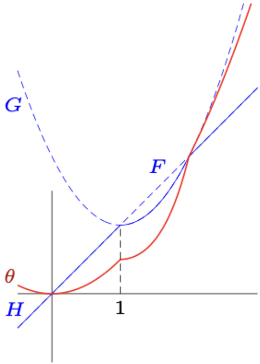
1) *The sequence $(\theta(x_k))_k$ decreases thus converges.*
2) *For any subsequence such that $(F'(x_k), G'(x_k), \lambda_k S_k)$ is bounded, $g_k \to 0$*

But no info on the type of the limit point.

Example of an undesirable limit point

Consider a simple example with
$n = 1$, $F(x) = x$,
$G(x) = 1 + (x-1)^2$, $x_0 = 3/2$.

For $x \in (1, 2)$, $F(x) \neq G(x)$.

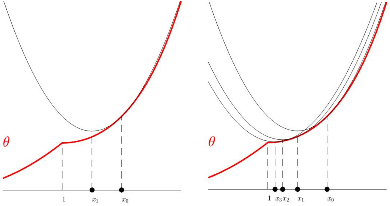


Counter-example

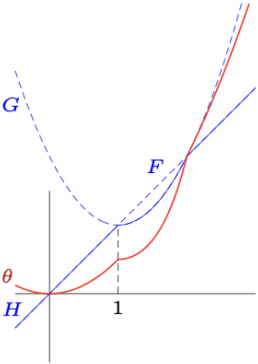
Example of an undesirable limit point

Consider a simple example with
$n = 1$, $F(x) = x$,
$G(x) = 1 + (x-1)^2$, $x_0 = 3/2$.

For $x \in (1,2)$, $F(x) \neq G(x)$.



First iterates, convergence to $x = 1$. The black curves are the quadratic models φ_x.



Counter-example

Conclusion

- $\ominus$ involved/heavy computations;
- $\ominus$ limited results
- $\oplus$ weak assumptions
- $\oplus$ improvements in sight ($\rightarrow\tau$)

Next: convergence with τ, understanding γ_+, γ_-, full algorithm...

Thanks for your attention! Any questions?
(or let's go enjoy the coffee break!)

Conclusion

- $\ominus$ involved/heavy computations;
- $\ominus$ limited results
- $\oplus$ weak assumptions
- $\oplus$ improvements in sight ($\rightarrow \tau$)

Next: convergence with τ, understanding γ_+, γ_-, full algorithm...

Thanks for your attention! Any questions?
(or let's go enjoy the coffee break!)

Bibliographic elements I

[CPS92] R.W. Cottle, J.-S. Pang, and R.E. Stone. *The Linear Complementarity Problem*. Academic Press, Boston, 1992.

[DFG19] J.-P. Dussault, M Frappier, and J.Ch. Gilbert. *Polyhedral Newton-min algorithms for complementarity problems*. Tech. rep. Inria Paris ; Université de Sherbrooke, 2019. DOI: hal-02306526.

[FP03] F. Facchinei and J.-S. Pang. *Finite-Dimensional Variational Inequalities and Complementarity Problems*. Springer Series in Operations Research. Springer, 2003.

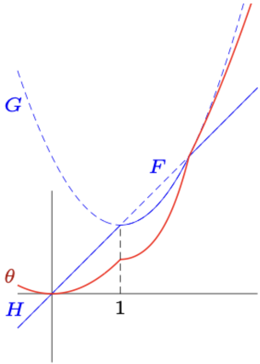
Outline

Details on the counter-example - 1

At point $x = 1$, $F(1) = 1 = G(1)$,
$\mathcal{E}^-(x) = \varnothing, \mathcal{E}^{0+}(x) = \{1\}$,
$F'(1) = 1, G'(1) = 0$
($x = 1$ is not regular)

$$g(\gamma_+) = \gamma_+ \times \underbrace{1}_{F'(1)} + (1 - \gamma_+) \times \underbrace{0}_{G'(1)} = \gamma_+$$

descent towards $x \in [0, 1]$ if $\gamma_+ > 0$



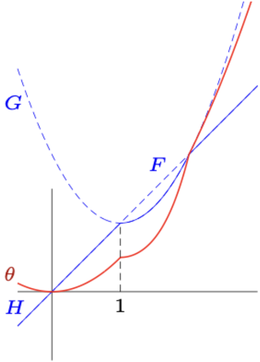
Counter-example

Details on the counter-example - 1

At point $x = 1$, $F(1) = 1 = G(1)$,
$\mathcal{E}^{-}(x) = \varnothing, \mathcal{E}^{0+}(x) = \{1\}$,
$F'(1) = 1, G'(1) = 0$
($x = 1$ is not regular)

$$g(\gamma_+) = \gamma_+ \times \underbrace{1}_{F'(1)} + (1 - \gamma_+) \times \underbrace{0}_{G'(1)} = \gamma_+$$

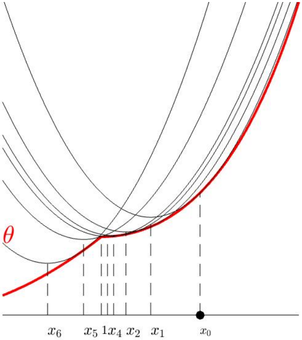
descent towards $x \in [0, 1]$ if $\gamma_+ > 0$



Counter-example

Details on the counter-example - 2

But if $x^k > 1$ for all k, the framework is not used.
With some simple framework for τ,



Usefulness of the lemma - 1

Consider the following example in $\mathbb{R}^2$, $x = [1; 1]$, for $\delta > 0$ "small"

$$F_1(x) = x_1 - 2 \qquad G_1(x) = 1 - 2x_1$$
$$F_2(x) = x_2 - 1 - \delta \qquad G_2(x) = 2x_2 - 2 - \delta$$

Clearly, $F_1(x) = -1 = G_1(x), F_2(x) = -\delta = G_2(x), \mathcal{E}^-(x) = \{1, 2\}$

$$g(\gamma) = (-1)[\gamma_1 e_1 - 2\overline{\gamma}_1 e_1] + (-\delta)[\gamma_2 e_2 + 2\overline{\gamma}_2 e_2].$$

For instance, $g(1/2, 1/2) = [1/2; -3/2]$, and

$$\begin{aligned}\theta'(x; -g) &= (-1)\min((-g)_1, -2(-g)_1) - \delta \min((-g)_2, 2(-g)_2) \\ &= -\min(-1/2, 1) - \delta \min(3\delta/2, 6\delta/2) \\ &= \frac{1}{2} - \frac{3}{2}\delta^2 > 0\end{aligned}$$

Usefulness of the lemma - 1

Consider the following example in $\mathbb{R}^2$, $x = [1; 1]$, for $\delta > 0$ "small"

$$F_1(x) = x_1 - 2 \qquad G_1(x) = 1 - 2x_1$$
$$F_2(x) = x_2 - 1 - \delta \qquad G_2(x) = 2x_2 - 2 - \delta$$

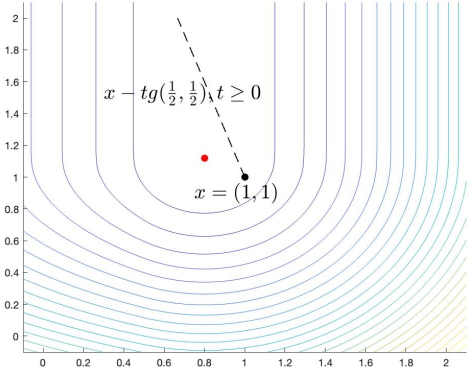
Clearly, $F_1(x) = -1 = G_1(x), F_2(x) = -\delta = G_2(x), \mathcal{E}^-(x) = \{1, 2\}$

$$g(\gamma) = (-1)[\gamma_1 e_1 - 2\overline{\gamma}_1 e_1] + (-\delta)[\gamma_2 e_2 + 2\overline{\gamma}_2 e_2].$$

For instance, $g(1/2, 1/2) = [1/2; -3/2]$, and

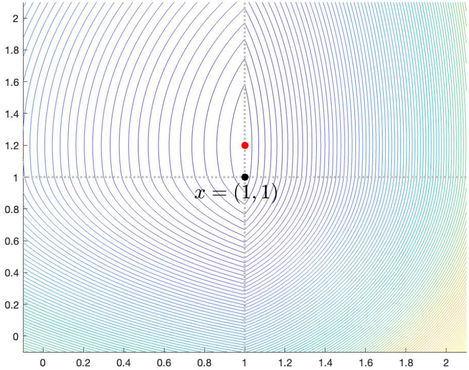
$$\begin{aligned}\theta'(x; -g) &= (-1)\min((-g)_1, -2(-g)_1) - \delta \min((-g)_2, 2(-g)_2) \\ &= -\min(-1/2, 1) - \delta \min(3\delta/2, 6\delta/2) \\ &= \frac{1}{2} - \frac{3}{2}\delta^2 > 0\end{aligned}$$

Usefulness of the lemma - 2



Level sets of φ_x with arbitrary weights $\gamma = (1/2, 1/2)$.

Usefulness of the lemma - 3



Function θ. The grey dotted lines are the kinks.

Usefulness of the lemma - 4

Observe that $\gamma_1 = 2/3, \gamma_2 = 1$ solves the lemma:

$$g(2/3, 1) = (-1)\left[\frac{2}{3} - 2\left(1 - \frac{2}{3}\right)\right] e_1 - \delta[1 + 2(1-1)]e_2 = -\delta e_2$$

and the lemma's equation reads

$$\begin{cases} \frac{2}{3}e_1^T(-g) + \frac{1}{3}(-2e_1)^T(-g) & = \min(e_1^T(-g), (-2e_1)^T(-g)) \\ 1 \times e_2^T(-g) + 0 \times (2e_2)^T(-g) & = \min(e_2^T(-g), (2e_2)^T(-g)) \end{cases}$$

which simplifies into

$$\begin{aligned} \frac{2}{3} \times 0 + \frac{1}{3} \times 0 & = \min(0, 0) \\ \delta & = \min(\delta, 2\delta) \end{aligned}$$

Usefulness of the lemma - 4

Observe that $\gamma_1 = 2/3, \gamma_2 = 1$ solves the lemma:

$$g(2/3, 1) = (-1)\left[\frac{2}{3} - 2\left(1 - \frac{2}{3}\right)\right] e_1 - \delta[1 + 2(1-1)]e_2 = -\delta e_2$$

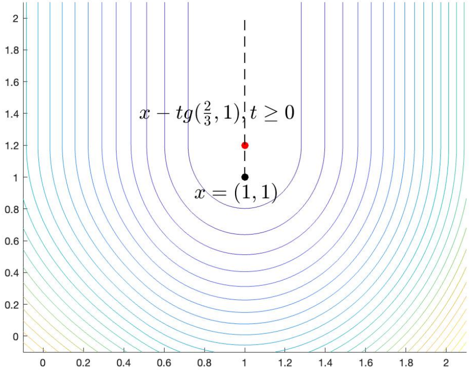
and the lemma's equation reads

$$\begin{cases} \frac{2}{3} e_1^T(-g) + \frac{1}{3}(-2e_1)^T(-g) & = \min(e_1^T(-g), (-2e_1)^T(-g)) \\ 1 \times e_2^T(-g) + 0 \times (2e_2)^T(-g) & = \min(e_2^T(-g), (2e_2)^T(-g)) \end{cases}$$

which simplifies into

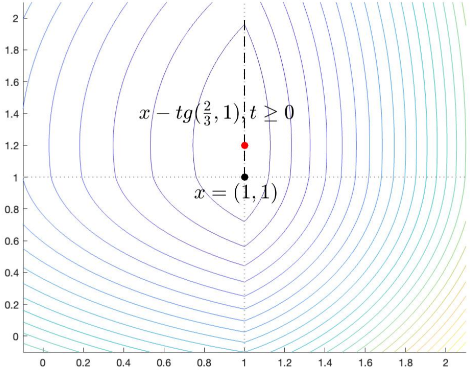
$$\begin{aligned} \tfrac{2}{3} \times 0 + \tfrac{1}{3} \times 0 & = \min(0, 0) \\ \delta & = \min(\delta, 2\delta) \end{aligned}$$

Usefulness of the lemma - 5



Level sets of φ_x with the weight given by the lemma.

Usefulness of the lemma - 6



Function θ. The grey dotted lines are the kinks.

Obtaining the weights - main idea

Goal

Finding a γ_+ such that $g(\gamma_+, \gamma_-(\gamma_+)) \neq 0$ (or stationarity).

$$\simeq \max_{\gamma_+ \in [0,1]^{\mathcal{E}^{0+}(x)}} \min_{\gamma_- \in [0,1]^{\mathcal{E}^{-}(x)}} ||g(\gamma_+, \gamma_-)||^2/2$$

where $g(\gamma_+, \gamma_-) = g_0 + M_+\gamma_+ + M_-\gamma_-$.

The outer max is for a convex function on a hypercube: combinatorial nature
(so $\rightsquigarrow \{0,1\}^{\mathcal{E}^{0+}(x)}$ and partitions).

Obtaining the weights - main idea

Goal

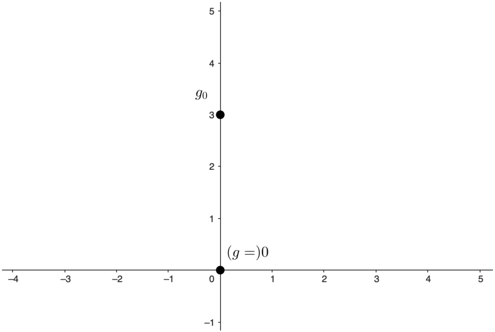
Finding a γ_+ such that $g(\gamma_+, \gamma_-(\gamma_+)) \neq 0$ (or stationarity).

$$\simeq \max_{\gamma_+ \in [0,1]^{\mathcal{E}^{0+}(x)}} \min_{\gamma_- \in [0,1]^{\mathcal{E}^{-}(x)}} ||g(\gamma_+, \gamma_-)||^2/2$$

where $g(\gamma_+, \gamma_-) = g_0 + M_+\gamma_+ + M_-\gamma_-$.

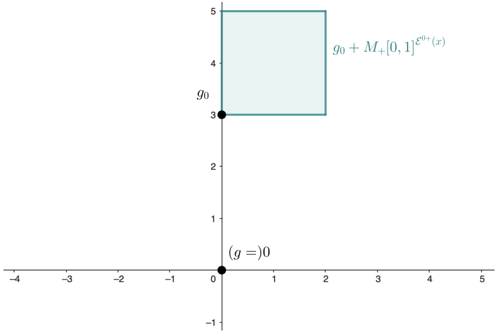
The outer max is for a convex function on a hypercube: combinatorial nature
(so $\rightsquigarrow \{0, 1\}^{\mathcal{E}^{0+}(x)}$ and partitions).

Obtaining the weights - illustration



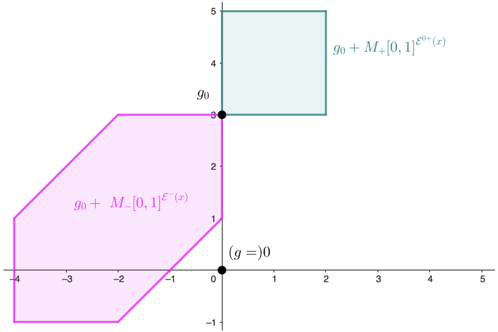
Constant term g_0. "Range" when adding only the γ_+. Or only the γ_-. "Range" with both. Specific γ_+ such that $g(\gamma_+, \gamma_-) = 0$. Illustration of $g(\gamma_+, \gamma_-) = 0$ being reachable.

Obtaining the weights - illustration



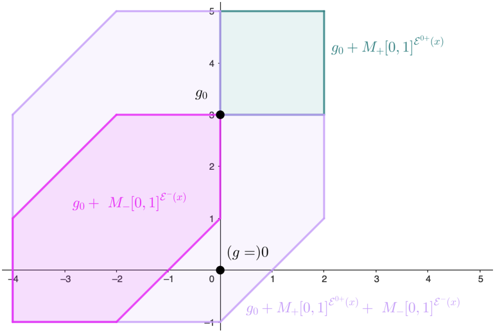
Constant term g_0. "Range" when adding only the γ_+. Or only the γ_-. "Range" with both. Specific γ_+ such that $g(\gamma_+, \gamma_-) = 0$. Illustration of $g(\gamma_+, \gamma_-) = 0$ being reachable.

Obtaining the weights - illustration



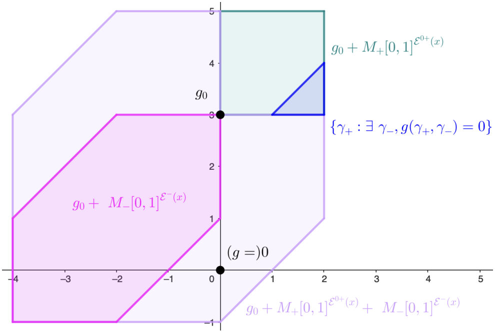
Constant term g_0. "Range" when adding only the γ_+. Or only the γ_-. "Range" with both. Specific γ_+ such that $g(\gamma_+, \gamma_-) = 0$. Illustration of $g(\gamma_+, \gamma_-) = 0$ being reachable.

Obtaining the weights - illustration



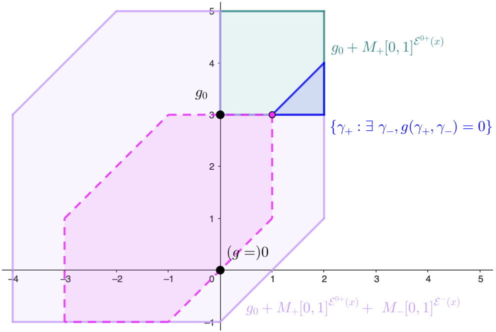
Constant term g_0. "Range" when adding only the γ_+. Or only the γ_-. "Range" with both. Specific γ_+ such that $g(\gamma_+, \gamma_-) = 0$. Illustration of $g(\gamma_+, \gamma_-) = 0$ being reachable.

Obtaining the weights - illustration



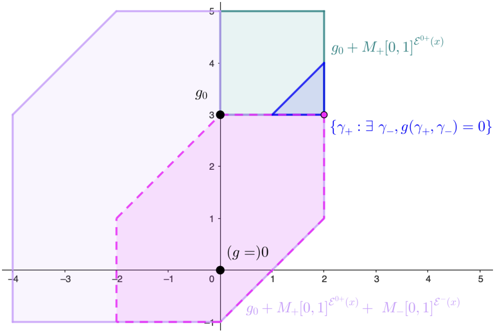
Constant term g_0. "Range" when adding only the γ_+. Or only the γ_-. "Range" with both. Specific γ_+ such that $g(\gamma_+, \gamma_-) = 0$. Illustration of $g(\gamma_+, \gamma_-) = 0$ being reachable.

Obtaining the weights - illustration



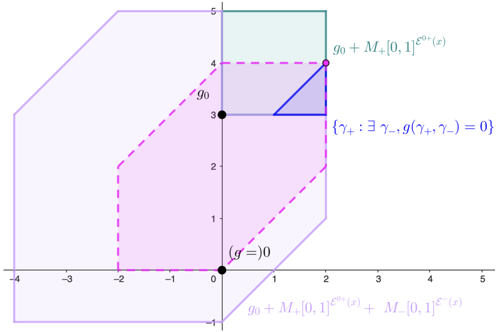
Constant term g_0. "Range" when adding only the γ_+. Or only the γ_-. "Range" with both. Specific γ_+ such that $g(\gamma_+, \gamma_-) = 0$. Illustration of $g(\gamma_+, \gamma_-) = 0$ being reachable.

Obtaining the weights - illustration



Constant term g_0. "Range" when adding only the γ_+. Or only the γ_-. "Range" with both. Specific γ_+ such that $g(\gamma_+, \gamma_-) = 0$. Illustration of $g(\gamma_+, \gamma_-) = 0$ being reachable.

Obtaining the weights - illustration



Constant term g_0. "Range" when adding only the γ_+. Or only the γ_-. "Range" with both. Specific γ_+ such that $g(\gamma_+, \gamma_-) = 0$. Illustration of $g(\gamma_+, \gamma_-) = 0$ being reachable.

Obtaining the weights - projection framework

Observe that the lemma's equation reads, with $g = g_1 + M_- \gamma_-$ and $M_- = (\nabla F_{\mathcal{E}-(x)} - \nabla G_{\mathcal{E}-(x)})\mathrm{Diag}(H_{\mathcal{E}-(x)})$

$$-\Gamma_- F'_{\mathcal{E}-(x)} g - \overline{\Gamma}_- G'_{\mathcal{E}-(x)} g = \min(-F'_{\mathcal{E}-(x)} g, -G'_{\mathcal{E}-(x)} g)$$

$$\Gamma_- F'_{\mathcal{E}-(x)} g + \overline{\Gamma}_- G'_{\mathcal{E}-(x)} g = \max(F'_{\mathcal{E}-(x)} g, G'_{\mathcal{E}-(x)} g)$$

$$\Gamma_- (F'_{\mathcal{E}-(x)} - G'_{\mathcal{E}-(x)}) g = \max((F'_{\mathcal{E}-(x)} - G'_{\mathcal{E}-(x)}) g, 0)$$

$$\Gamma_- M_-^\top g = \min(M_-^\top g, 0)$$

$$\begin{cases} (M_-^\top g)_i > 0 & \text{and } \gamma_i = 0 \\ (M_-^\top g)_i < 0 & \text{and } \gamma_i = 1 \\ (M_-^\top g)_i = 0 & \text{and } \gamma_i \in [0,1] \end{cases}$$

This is a reformulation of $\mathrm{P}_{M_-[0,1]^{\mathcal{E}-(x)}}(g_0 + M_+ \gamma_+)$

Obtaining the weights - projection framework

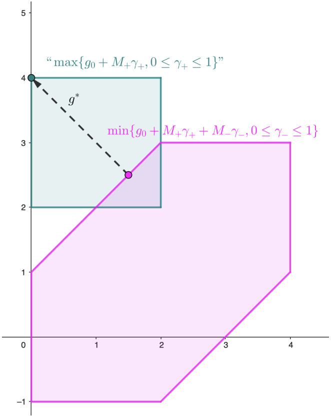
Observe that the lemma's equation reads, with $g = g_1 + M_-\gamma_-$ and $M_- = (\nabla F_{\mathcal{E}^-(x)} - \nabla G_{\mathcal{E}^-(x)})\mathrm{Diag}(H_{\mathcal{E}^-(x)})$

$$-\Gamma_- F'_{\mathcal{E}^-(x)} g - \overline{\Gamma}_- G'_{\mathcal{E}^-(x)} g = \min(-F'_{\mathcal{E}^-(x)} g, -G'_{\mathcal{E}^-(x)} g)$$
$$\Gamma_- F'_{\mathcal{E}^-(x)} g + \overline{\Gamma}_- G'_{\mathcal{E}^-(x)} g = \max(F'_{\mathcal{E}^-(x)} g, G'_{\mathcal{E}^-(x)} g)$$
$$\Gamma_-(F'_{\mathcal{E}^-(x)} - G'_{\mathcal{E}^-(x)}) g = \max((F'_{\mathcal{E}^-(x)} - G'_{\mathcal{E}^-(x)}) g, 0)$$
$$\Gamma_- M_-^{\mathsf{T}} g = \min(M_-^{\mathsf{T}} g, 0)$$

$$\begin{cases} (M_-^{\mathsf{T}} g)_i > 0 & \text{and } \gamma_i = 0 \\ (M_-^{\mathsf{T}} g)_i < 0 & \text{and } \gamma_i = 1 \\ (M_-^{\mathsf{T}} g)_i = 0 & \text{and } \gamma_i \in [0,1] \end{cases}$$

This is a reformulation of $\mathrm{P}_{M_-[0,1]^{\mathcal{E}^-(x)}}(g_0 + M_+\gamma_+)$

Obtaining the weights - projection illustrated



Projection (after a change of variables). The top left teal point is the furthest from the magenta zone. (Zonotope/combinatorial geometry)