Contributions aux méthodes numériques pour les problèmes de complémentarité et problèmes d'optimisation sous contraintes de complémentarité

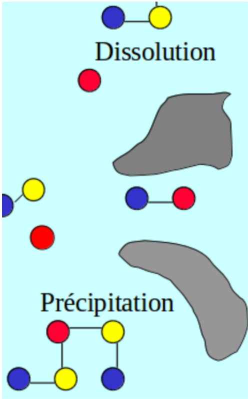
Tangi Migot

Soutenance de thèse - 06 octobre 2017

INSA Institut National des Sciences Appliquées Rennes
IRMAR
Université Bretagne Loire

Motivation: Equilibrium Problems in Geochemistry

Precipitation-dissolution reactions in geochemistry



p: concentration of a mineral,
c: concentration of aqueous components.

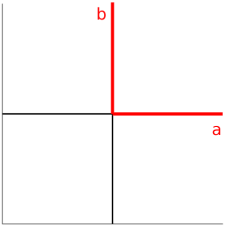
Action-Mass Law
2 possibles states (solid or liquid):

1. $p = 0, \ K_p - \gamma(c) \geq 0;$
2. $p \geq 0, \ K_p - \gamma(c) = 0;$

Motivation: The Complementarity Problem (CP)

Consider the following set of constraints:

$$C = \{(a, b) \in \mathbb{R}^q \times \mathbb{R}^q \mid 0 \leq a \perp b \geq 0\}.$$



1. In general, $a \equiv G(x)$ and $b \equiv H(x)$ with two maps $G, H : \mathbb{R}^n \to \mathbb{R}^q$;
2. Even in the "most simple" case with G and H affine the problem of finding a "feasible" point in C is NP-hard in general.

Motivation: Non-linear Programming

Consider a non-linear program with an objective function $f : \mathbb{R}^n \to \mathbb{R}$ and constraints $g : \mathbb{R}^n \to \mathbb{R}^p$, $h : \mathbb{R}^n \to \mathbb{R}^m$ so that

$$\min_{x \in \mathbb{R}^n} f(x) \text{ s.t. } g(x) \leq 0, h(x) = 0. \tag{NLP}$$

Motivation: Non-linear Programming

Consider a non-linear program with an objective function $f : \mathbb{R}^n \to \mathbb{R}$ and constraints $g : \mathbb{R}^n \to \mathbb{R}^p$, $h : \mathbb{R}^n \to \mathbb{R}^m$ so that

$$\min_{x \in \mathbb{R}^n} f(x) \text{ s.t. } g(x) \leq 0, h(x) = 0. \tag{NLP}$$

For any "qualified" local minimum (x^*) of (NLP), there exists a Lagrange multiplier $\lambda := (\lambda^g, \lambda^h)$ such that

$$\begin{aligned} -\nabla f(x^*) &= \sum_{i=1}^{m} \lambda_i^g \nabla g_i(x^*) + \sum_{i=1}^{p} \lambda_i^h \nabla h_i(x^*), \\ h(x^*) &= 0, \; 0 \leq -g(x^*) \perp \lambda^g \geq 0. \end{aligned} \tag{KKT}$$

Motivation: Non-linear Programming

Consider a non-linear program with an objective function $f : \mathbb{R}^n \to \mathbb{R}$ and constraints $g : \mathbb{R}^n \to \mathbb{R}^p$, $h : \mathbb{R}^n \to \mathbb{R}^m$ so that

$$\min_{x \in \mathbb{R}^n} f(x) \text{ s.t. } g(x) \leq 0, h(x) = 0. \tag{NLP}$$

For any "qualified" local minimum (x^*) of (NLP), there exists a Lagrange multiplier $\lambda := (\lambda^g, \lambda^h)$ such that

$$\begin{aligned} -\nabla f(x^*) = \sum_{i=1}^{m} \lambda_i^g \nabla g_i(x^*) + \sum_{i=1}^{p} \lambda_i^h \nabla h_i(x^*), \\ h(x^*) = 0,\ 0 \leq -g(x^*) \perp \lambda^g \geq 0. \end{aligned} \tag{KKT}$$

Application of CP

The KKT conditions form a complementarity problem.

Motivation: Bilevel Programming

In many applications the scientist/engineer/leader has to optimize depending on the answer of other people (= another optimization problem in the constraints).

$$\begin{aligned} \min_{x,y \in \mathbb{R}^{n_0} \times \mathbb{R}^{n_1}} \ & f_0(x,y) \\ \text{s.t. } & g_0(x,y) \leq 0, h_0(x,y) = 0, \\ & y \in S(x), \end{aligned} \tag{BP}$$

where

$$S(x) = \arg\min_{y \in \mathbb{R}^{n_1}} \{ f_1(x,y) \text{ s.à. } g_1(y) \leq 0, h_1(y) = 0 \}.$$

Motivation: Bilevel Programming

In many applications the scientist/engineer/leader has to optimize depending on the answer of other people (= another optimization problem in the constraints).

$$\begin{aligned} \min_{x,y\in\mathbb{R}^{n_0}\times\mathbb{R}^{n_1}} \; & f_0(x,y) \\ \text{s.t. } \; & g_0(x,y)\leq 0, h_0(x,y)=0, \\ & y\in S(x), \end{aligned} \tag{BP}$$

where

$$S(x)=\arg\min_{y\in\mathbb{R}^{n_1}}\{f_1(x,y) \text{ s.à. } g_1(y)\leq 0, h_1(y)=0\}.$$

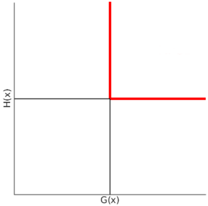
Optimistic Bilevel Program

Replace $S(x)$ by its optimality conditions, we optimize a function over a complementarity set. We call the resulting problem a **Mathematical Program with Complementarity Constraints**.

The Complementarity Set

Consider the following set of constraints:

$$C = \{(a, b) \in \mathbb{R}^q \times \mathbb{R}^q \mid 0 \leq a \perp b \geq 0\}.$$

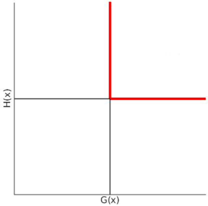


1. non-convex domain with "kink";
2. in general, $a \equiv G(x)$ and $b \equiv H(x)$ with two maps $G, H : \mathbb{R}^n \to \mathbb{R}^q$ -> non-connected domain;
3. thin domain (i.e. $\nexists x^* \in \mathbb{R}^n,\ G(x^*) > 0, H(x^*) > 0$).

The Complementarity Set

Consider the following set of constraints:

$$C = \{(a, b) \in \mathbb{R}^q \times \mathbb{R}^q \mid 0 \leq a \perp b \geq 0\}.$$



1. non-convex domain with "kink";
2. in general, $a \equiv G(x)$ and $b \equiv H(x)$ with two maps $G, H : \mathbb{R}^n \to \mathbb{R}^q$ -> non-connected domain;
3. thin domain (i.e. $\nexists x^* \in \mathbb{R}^n,\ G(x^*) > 0, H(x^*) > 0$).

Natural idea:

regularization or **relaxation** of the domain.

The θ Regularization

The θ's function

For a regularization parameter $r > 0$, we consider for $x \in \mathbb{R}_+$

$$\theta_r(x) \approx \|x\|_0,$$

where for $z \in \mathbb{R}^n$, $\|z\|_0 := \#\{z_i \neq 0\}$.

In this case, the complementarity can be "approximated" with

$$a \perp b \approx \theta_r(a) + \theta_r(b) \leq 1.$$

The θ Regularization

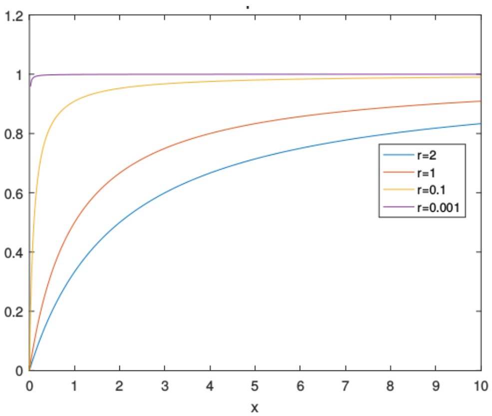
Given $r > 0$. Let $\theta_r : \mathbb{R} \to]-\infty, 1]$ be a smooth (C^2 or C^1), non-decreasing, concave function such that

1. $\theta_r(0) = 0$;
2. $\lim_{\frac{x}{r} \to \infty} \theta_r(x) = 1$;
3. $\theta_r(x) < 0$ for $x < 0$.

These properties yields to

$$\lim_{r \to 0^+} \theta_r(x) = \begin{cases} 1, & \text{if } x > 0, \\ 0, & \text{otherwise.} \end{cases}$$

The θ Regularization



Examples for $x \geq 0$

$$\theta_r^1(x) = \frac{x}{x+r} \text{ and } \theta_r^2(x) = 1 - \exp(-\frac{x}{r})$$

Outline

Our aim is to derive fast and efficient algorithms, so our classical framework is composed of:

1. continuously differentiable data;
2. computation of stationary point (or at best local optima).

Outline

Outline

The Problem of Sparse Optimization

Find $x \in \mathbb{R}^n$ the sparsest solution over a polyhedron P:

$$\min_{x \in P} \|x\|_0. \qquad (P_0)$$

$\emptyset \neq P = \{x \in \mathbb{R}^n | \ b \in \mathbb{R}^m, \ Ax \leq b\} \cap \mathbb{R}^n_+$ (many results are valid for a convex set $P \subset \mathbb{R}^n$).

Many popular applications

Compressed sensing, image recovery,...

A General Family of Concave Functions

We consider for $r > 0$ the following problem

$$\min_{x \in P} \sum_{i=1}^{n} \theta_r(x_i) = \min_{x \in P} \Theta_r(x).$$

1. Concave optimization problem;
2. By definition of θ_r, it holds

$$\lim_{r \to 0^+} \Theta_r(x) = \|x\|_0;$$

3. **Existence of solution**, whenever $P \subset \mathbb{R}^n_+$ is non-empty, convex and closed, results from asymptotic analysis.

An Homotopy Method

$$\min_{x \in P} ||x||_1 \;\rightarrow\; \min_{x \in P} \Theta_r(x) \;\rightarrow\; \min_{x \in P} ||x||_0$$

An Homotopy Method

$$\min_{x \in P} ||x||_1 \rightarrow \min_{x \in P} \Theta_r(x) \rightarrow \min_{x \in P} ||x||_0$$

We get an homotopy technique that should improve the classical convex approximation.

$(P_1) \rightarrow (P_r)$

Taylor theorem in one dimension and $\theta_r(x) := \theta(x/r)$ yields to

$$\theta(x/r) = \frac{x}{r}\theta'(0) + o(x/r).$$

As $r > 0$, we can use a scaling technique

$$\min_{x \in P} \Theta_r(x) \iff \min_{x \in P} r\Theta_r(x).$$

A Sufficient Convergence Condition

- $k = ||x^*||_0 < n$ be the (unknown) optimal value of problem (P_0);
- $S^*_{||.||_0}$ the set of solutions of (P_0);
- $x_r \in S^*_r$;
- θ functions where $\theta \geq \theta^1$;

Theorem (Exact Penalization, *Haddou-Migot,15'*)

$$\theta_r\left(\min_{(x_r)_i \neq 0} (x_r)_i\right) \geq \frac{k}{k+1} \Longrightarrow x_r \in S^*_{||.||_0}.$$

We can bound $\frac{k}{k+1}$ by $\frac{||x_0||_0}{||x_0||_0+1}$, which is known.

A Sufficient Convergence Condition

- $k = ||x^*||_0 < n$ be the (unknown) optimal value of problem (P_0);
- $S^*_{||.||_0}$ the set of solutions of (P_0);
- $x_r \in S^*_r$;
- θ functions where $\theta \geq \theta^1$;

Theorem (Exact Penalization, *Haddou-Migot,15'*)

$$\theta_r\left(\min_{(x_r)_i \neq 0} (x_r)_i\right) \geq \frac{k}{k+1} \Longrightarrow x_r \in S^*_{||.||_0}.$$

We can bound $\frac{k}{k+1}$ by $\frac{||x_0||_0}{||x_0||_0+1}$, which is known.

Numerics on random test problems

The θ regularization manages to improve the solution provided by the convex ℓ_1 problem.

Outline

The Absolute Value Equation (AVE)

AVE consists in finding $x \in \mathbb{R}^n$ that verifies

$$Ax - |x| = b,$$

with $A \in \mathbb{R}^{n \times n}$ and $b \in \mathbb{R}^n$.

1. Application: ODE with absolute values;
2. Difficulties in presence of degeneracy (some singular values of A are 1). Mangasarian [2007 - 2014] proposes bilinear or concave reformulations;
3. Reformulation with complementarity constraints of the absolute value:

$$|x| = x^+ + x^-, \; 0 \leq x^+ \perp x^- \geq 0 \Longrightarrow x = x^+ - x^-.$$

θ Regularization of AVE

$$
\begin{aligned}
\min_{x^+,x^-} \quad & \Theta_r(x^+) + \Theta_r(x^-) \\
\text{s.t.} \quad & |A(x^+ - x^-) - (x^+ + x^-) - b| \leq g(r)(|A| + I)e, \\
& x^+ \geq 0, \ \ x^- \geq 0, \\
& x^+ + x^- \geq g(r),
\end{aligned}
$$

where $r = o(g(r))$ (for instance $g(r) = r^\alpha$ with $0 < \alpha < 1$).

Remark

The constraint $x^+ + x^- \geq g(r)$ avoid a compensation phenomenon in the objective function.

θ Regularization of AVE: Theory

1. Algorithm: Homotopy technique for $\{r\}$ with $r \to 0^+$;

θ Regularization of AVE: Theory

1. Algorithm: Homotopy technique for $\{r\}$ with $r \to 0^+$;
2. Error bound:

Theorem (*Abdallah-Haddou-Migot,18'*)

Let $\{x^{r+}, x^{r-}\} \to (\bar{x}^+, \bar{x}^-)$. *Then,*

$$d_{S^*_{(AVE)}}(x^{r+} - x^{r-}) = O(g(r)),$$

where $d_{S^*_{(AVE)}}$ *denotes the distance (2-norm) to the set of solutions.*

θ Regularization of AVE: Numerics

We compare 4 methods tailored for general AVE:

- TAVE method (θ regularization using SLA);
- TAVE2 which is the same algorithm with the different objective

$$\sum_{i=1}^{n} \theta_r(x_i^+) + \theta_r(x_i^-) - \theta_r(x_i^+ + x_i^-);$$

- concave minimization method CMM from [Mangasarian, 07'];
- successive linear programming method LPM from [Mangasarian, 14'].

θ Regularization of AVE: Numerics

We compare 4 methods tailored for general AVE:

- TAVE method (θ regularization using SLA);
- TAVE2 which is the same algorithm with the different objective

$$\sum_{i=1}^{n} \theta_r(x_i^+) + \theta_r(x_i^-) - \theta_r(x_i^+ + x_i^-);$$

- concave minimization method CMM from [Mangasarian, 07'];
- successive linear programming method LPM from [Mangasarian, 14'].

Numerical results on random problems

TAVE significantly reduces the number of unsolved problems.

θ Regularization of AVE: Numerics

We compare 4 methods tailored for general AVE:

- TAVE method (θ regularization using SLA);
- TAVE2 which is the same algorithm with the different objective

$$\sum_{i=1}^{n} \theta_r(x_i^+) + \theta_r(x_i^-) - \theta_r(x_i^+ + x_i^-);$$

- concave minimization method CMM from [Mangasarian, 07'];
- successive linear programming method LPM from [Mangasarian, 14'].

Numerical results on random problems

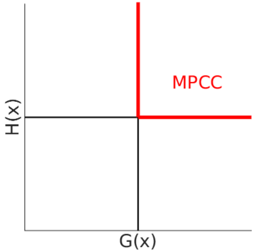
TAVE significantly reduces the number of unsolved problems.
Perspectives: TAVE2 is doing even better.

Outline

The Mathematical Program with Complementarity Constraints (MPCC)

Let f, h, g, G, H be continuously differentiable maps.

$$\begin{aligned} &\min_{x \in \mathbb{R}^n} f(x) \\ &\text{s.t. } \; h(x) = 0, \; g(x) \le 0, \\ &\qquad 0 \le G(x) \perp H(x) \ge 0, \end{aligned} \tag{MPCC}$$



Feasible set of $0 \le G(x) \perp H(x) \ge 0$

Major difficulty :

Classical CQs, fail to hold in general $\Longrightarrow$ no KKT.

Example

$$\min_{x\in\mathbb{R}^2} x_1^2 + x_2^2 - x_3$$

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

Obviously the point $(0,0,0)^T$ *is the global minimum. There exists multipliers* $\lambda^{g_1}, \lambda^{g_2}, \lambda^G, \lambda^H, \lambda^{\perp} = (1,0,-4,0,0)$ *but none with the correct signs regarding the KKT conditions.*

Major difficulty :

Classical CQs, fail to hold in general $\Longrightarrow$ no KKT.

Example

$$\min_{x\in\mathbb{R}^2} x_1^2 + x_2^2 - x_3$$
$$\begin{aligned} s.t. \quad & -4x_1 + x_3 \le 0, \\ & -4x_2 + x_3 \le 0, \\ & 0 \le x_1 \perp x_2 \ge 0 \end{aligned}$$

Obviously the point $(0,0,0)^T$ *is the global minimum. There exists multipliers* $\lambda^{g_1}, \lambda^{g_2}, \lambda^G, \lambda^H, \lambda^\perp = (1, 0, -4, 0, 0)$ *but none with the correct signs regarding the KKT conditions.*

What is a stationary point in the MPCC sense ?

MPCC-Lagrangian function of (MPCC) as

$$\mathcal{L}_{MPCC}(x,\lambda) = f(x) + g(x)^T\lambda^g + h(x)^T\lambda^h - G(x)^T\lambda^G - H(x)^T\lambda^H,$$

$$\mathcal{I}^{00} := \{i \mid G_i(x) = 0, H_i(x) = 0\},$$
$$\mathcal{I}^{+0} := \{i \mid G_i(x) > 0, H_i(x) = 0\},$$
$$\mathcal{I}^{0+} := \{i \mid G_i(x) = 0, H_i(x) > 0\}.$$

Definition

x^* feasible for (MPCC) is said

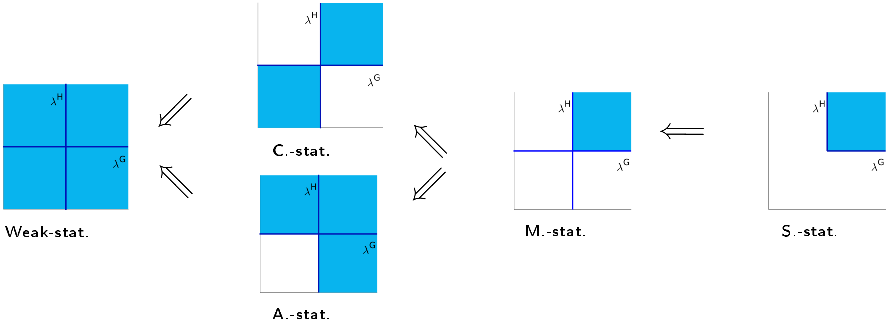
- *Weak-stationary if there exists $\lambda = (\lambda^g, \lambda^h, \lambda^G, \lambda^H) \in \mathbb{R}^{p+q+2m}$ such that*

$$\nabla_x \mathcal{L}_{MPCC}(x^*, \lambda^g, \lambda^h, \lambda^G, \lambda^H) = 0,$$
$$\lambda^g_{\mathcal{I}_g} \geq 0, \ \lambda^G_{\mathcal{I}^{+0}} = 0, \ \lambda^H_{\mathcal{I}^{0+}} = 0.$$

Moreover, x^* weak-stationary is said

- C.-stationary: $\lambda_i^G \lambda_i^H \geq 0$;
- A.-stationary: $\lambda_i^G \geq 0$ or $\lambda_i^H \geq 0$;
- M.-stationary: either $\lambda_i^G > 0$, $\lambda_i^H > 0$ or $\lambda_i^G \lambda_i^H = 0$;
- S.-stationary: $\lambda_i^G \geq 0$, $\lambda_i^H \geq 0$.

For all $i \in \mathcal{I}^{00} := \{i \mid G_i(x^*) = H_i(x^*) = 0\}$.



Necessary Optimality Conditions for MPCC

Theorem (Flegel-Kanzow, 06')

A local minimum of (MPCC) that satisfies MPCC-GCQ or any stronger MPCC-CQ is an M-stationary point.

- A classical KKT-point is an S-starionary point.
- We will not get into the details of MPCC-CQs here.

Theorem (Flegel-Kanzow, 06')

A local minimum of (MPCC) that satisfies MPCC-GCQ or any stronger MPCC-CQ is an M-stationary point.

- A classical KKT-point is an S-starionary point.
- We will not get into the details of MPCC-CQs here.

Goal/Motivation :

- Numerical methods should converge to M-stationary points

Relax the Constraint : $0 \leq G(x) \perp H(x) \geq 0$

- $+$: Improved regularity ($=$ satisfy a CQ)
- $-$: Convergence properties ?

$$\begin{aligned} &\min_{x \in \mathbb{R}^n} f(x) \\ &\text{s.t } h(x) = 0, \; g(x) \leq 0, \\ &\qquad G(x) \geq -\bar{t}, \; H(x) \geq -\bar{t}, \\ &\qquad \color{red}{\Phi(G(x), H(x); t) \leq 0.} \end{aligned} \qquad (Relax_{t,\bar{t}})$$

$t, \bar{t}$ are positive parameters.

Generic Regularization Algorithm

Data: x^0 an initial point, $(t_0, \bar{t}_0)$ initial parameters, $\sigma_t \in (0,1)$ parameters update;

Set $k := 0$, $(t_k, \bar{t}_k) := (t_0, \bar{t}_0)$;

repeat

$\quad (t_{k+1}, \bar{t}_{k+1}) = \sigma_t(t_k, \bar{t}_k)$;

$\quad x^{k+1} :=$ stationary point of $(Relax_{t,\bar{t}})$ with x^k initial point;

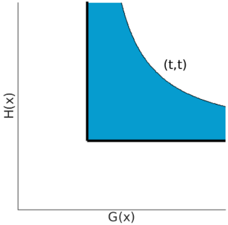
$\quad k := k + 1$;

until $\underline{x^{k+1}\text{ is "M-stationary of MPCC"}}$;

Relax the constraint: $0 \leq G(x) \perp H(x) \geq 0$

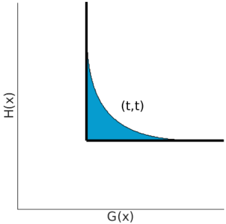
- $+$: Improved regularity ($=$ satisfy a CQ)
- $-$: Convergence properties ?

Relaxation methods that converge to C-stationary points ($t \downarrow 0$):



$$\theta_r(G(x)) + \theta_r(H(x)) \leq 1$$

or Scholtes, 2000 for θ^1.



Steffensen-Ulbrich, 2010.

A Unified Framework for Regularization Methods

Assume that the relaxation map $\Phi(G(x), H(x); t)$ (C^1) is of the form

$$\Phi(G(x), H(x); t) = 0 \iff F_G(G(x), H(x); t)F_H(G(x), H(x); t) = 0,$$

where

$$\begin{aligned} F_G(G(x), H(x); t) &= G(x) - \psi(H(x); t), \\ F_H(G(x), H(x); t) &= H(x) - \psi(G(x); t). \end{aligned}$$

A Unified Framework for Regularization Methods

Assume that the relaxation map $\Phi(G(x), H(x); t)$ (C^1) is of the form

$$\Phi(G(x), H(x); t) = 0 \iff F_G(G(x), H(x); t) F_H(G(x), H(x); t) = 0,$$

where

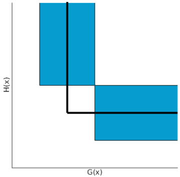
$$\begin{aligned} F_G(G(x), H(x); t) &= G(x) - \psi(H(x); t), \\ F_H(G(x), H(x); t) &= H(x) - \psi(G(x); t). \end{aligned}$$

Assumptions:

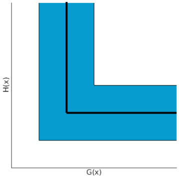
1. $\lim_{\|t\| \to 0} \|\psi(z; t)\| = 0 \ \forall z \in \mathbb{R}^q$;
2. $\lim_{\|t\| \to 0} \nabla_{G(x)} \Phi = H(x)$ and $\lim_{\|t\| \to 0} \nabla_{H(x)} \Phi = G(x)$;
3. $\lim_{\|t\| \to 0} \left. \frac{\partial \psi(x; t)}{\partial x} \right|_{x=z} = 0 \ \forall z \in \mathbb{R}^q$.

Relax the Constraint: $0 \leq G(x) \perp H(x) \geq 0$

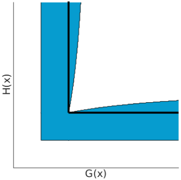
Regularization methods that belong to the **Unified Framework**:



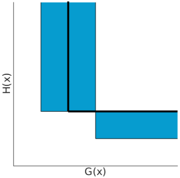
Kadrani-Dussault-Benchakroun,2009
$\psi(z;t) = t, \;\; t \in \mathbb{R}$



Kanzow-Schwartz,2013
$\psi(z;t) = t, \;\; t \in \mathbb{R}$



Dussault-Haddou-Migot,2016
"Butterfly" method
$\psi(z;t) = t_2\theta(z, t_1),$
$t \in \mathbb{R}^2$



Asymmetric regularization
$\psi(z;t)^1 = t$ and
$\psi(z;t)^2 = 0, \;\; t \in \mathbb{R}$

Convergence Theorem

1. $\{t_k, \bar{t}_k\} \searrow 0$;
2. $\{x^k, \lambda_k\}$ a sequence of stationary (KKT-) points of $(Relax_{t,\bar{t}})$ for all $k \in \mathbb{N}$ with $x^k \to x^*$;
3. Suitable MPCC-CQ [1] holds at x^*;

Theorem (*Dussault-Haddou-Kadrani-Migot,17'*)

x^* is an M-stationary point.

[1] MPCC-CRSC (or MPCC-CCP)

Numerical Comparison

- A sensitivity analysis on several values of the parameters, 7 values for t_0 and 3 for σ_t;
- MacMPEC is a collection of test problems from real-world applications available in AMPL,

Leyffer, Sven.
MacMPEC: AMPL collection of MPECs.
www.mcs.anl.gov/leyffer/MacMPEC, 2000.

- Simulations with three solvers SNOPT, IPOPT and MINOS.

Numerical Comparison

- A sensitivity analysis on several values of the parameters, 7 values for t_0 and 3 for σ_t;
- MacMPEC is a collection of test problems from real-world applications available in AMPL,

Leyffer, Sven.
MacMPEC: AMPL collection of MPECs.
www.mcs.anl.gov/leyffer/MacMPEC, 2000.

- Simulations with three solvers SNOPT, IPOPT and MINOS.

Results

The butterfly relaxation(s) give promising results.

Classical Definition of ϵ-Stationary Points

Remark

Practical implementation of the regularization method: at each step we compute an ϵ-**stationary point** and not an exact one.

Classical Definition of ϵ-Stationary Points

Remark

Practical implementation of the regularization method: at each step we compute an ϵ-**stationary point** and not an exact one.

Main problem : ϵ-stationary sequences may converge to weak-stationary points.

Christian Kanzow and Alexandra Schwartz.
The Price of Inexactness: Convergence Properties of Relaxation Methods for Mathematical Programs with Complementarity Constraints Revisited.
Mathematics of Operations Research, 40(2):253–275, may 2015.

Strong ϵ-Stationary Points

A new definition of approximate stationary point, so called **strong epsilon-stationary point**:

$$\left\| \nabla \mathcal{L}_R(x, \lambda^g, \lambda^h, \lambda^G, \lambda^H, \lambda^\Phi) \right\|_\infty \leq \epsilon_k$$

with

$$\|h(x)\|_\infty \leq \epsilon_k,\ g(x) \leq \epsilon_k,\ \lambda^g \geq 0,\ \|\lambda^g \circ g(x)\|_\infty \leq \epsilon_k,$$

$$G(x) + \bar{t}_k \geq -\epsilon_k,\ \lambda^G \geq 0,\ \|\lambda^G \circ (G(x) + \bar{t}_k)\|_\infty \leq \epsilon_k,$$

$$H(x) + \bar{t}_k \geq -\epsilon_k,\ \lambda^H \geq 0,\ \|\lambda^H \circ (H(x) + \bar{t}_k)\|_\infty \leq \epsilon_k,$$

$$\Phi(G(x^k), H(x^k); t_k) \leq \cancel{\epsilon_k}\ 0,\ \lambda^\Phi \geq 0,\ \|\lambda^\Phi \circ \Phi(G(x^k), H(x^k); t_k)\|_\infty \leq \cancel{\epsilon_k}\ 0.$$

$$\Uparrow \qquad\qquad \Uparrow$$

Strong ϵ-Convergence Theorem

1. $\{t_k, \bar{t}_k\} \searrow 0$;
2. $\epsilon_k = o(\bar{t}_k)$;
3. $\{x^k, \lambda_k\}$ a sequence of strong ϵ_k-stationary (KKT-) points of $(Relax_{t,\bar{t}})$ for all $k \in \mathbb{N}$ with $x^k \to x^*$;
4. Suitable MPCC-CQ [1] holds at x^*;

Theorem (*Dussault-Haddou-Kadrani-Migot,17'*)

x^* is an M-stationary point.

[1] MPCC-CRSC (or MPCC-CCP)

Strong ϵ-Convergence Theorem

1. $\{t_k, \bar{t}_k\} \searrow 0$;
2. $\epsilon_k = o(\bar{t}_k)$;
3. $\{x^k, \lambda_k\}$ a sequence of strong ϵ_k-stationary (KKT-) points of $(Relax_{t,\bar{t}})$ for all $k \in \mathbb{N}$ with $x^k \to x^*$;
4. Suitable MPCC-CQ [1] holds at x^*;

Theorem (*Dussault-Haddou-Kadrani-Migot,17'*)

x^* is an M-stationary point.

Question: Is it possible to design a method that computes strong ϵ-stationary point ?

[1] MPCC-CRSC (or MPCC-CCP)

Strong ϵ-Convergence Theorem

1. $\{t_k, \bar{t}_k\} \searrow 0$;
2. $\epsilon_k = o(\bar{t}_k)$;
3. $\{x^k, \lambda_k\}$ a sequence of strong ϵ_k-stationary (KKT-) points of $(Relax_{t,\bar{t}})$ for all $k \in \mathbb{N}$ with $x^k \to x^*$;
4. Suitable MPCC-CQ [1] holds at x^*;

Theorem (*Dussault-Haddou-Kadrani-Migot,17'*)

x^* is an M-stationary point.

Question: Is it possible to design a method that computes strong ϵ-stationary point ? $\boxed{\text{Yes !}}$

[1] MPCC-CRSC (or MPCC-CCP)

Formulation with slack variables

$$
\begin{aligned}
\min_{x\in\mathbb{R}^n, s_G\in\mathbb{R}^q, s_H\in\mathbb{R}^q} \quad & f(x) \\
\text{s.t} \quad & h(x)=0,\ \ g(x)\leq 0, \\
& G(x)=s_G,\ \ H(x)=s_H, \\
& s_G \geq -\bar{t},\ \ s_H \geq -\bar{t}, \\
& \Phi(s_G, s_H; t) \leq 0.
\end{aligned}
$$

Motivations of slack variables:

- Existence of strong ϵ-stationary points in a neighbourhood of an M-stationary point.
- Algorithmic computation of strong ϵ-stationary points.

Regularized-Penalized Problem with Slack Variables

$$\begin{aligned}
\min_{x, s_G, s_H} \quad & f(x) + \frac{1}{\rho}\left(\| \max(g(x), 0), h(x), G(x) - s_G, H(x) - s_H\|_2^2\right) \\
\text{s.t.} \quad & s_G \geq -\bar{t}, \;\; s_H \geq -\bar{t}, \\
& \Phi(s_G, s_H; t) \leq 0,
\end{aligned} \qquad (Slack_Pen_Relax_t)$$

Active-Set-Penalization for $(Relax_{t,\bar{t}})$

1. Project the initial point on the feasible set;
2. Let $\mathcal{W}(s; t, \bar{t})$ be the set of active constraints among the constraints

$$s_G \geq -\bar{t}, \;\; s_H \geq -\bar{t}, \;\; \Phi(s_G, s_H; t) \leq 0,$$

   where $\Phi(s_G, s_H; t) = (s_G - \psi(s_H; t))(s_H - \psi(s_G; t))$;
3. Minimize the unconstrained problem $(Slack\_Pen\_Relax_t)$
   - Compute the gradient in the working subspace using composition rule of the derivative;
   - Restricted step to remain feasible for the relaxed complementarity constraints;
4. Compute the Lagrange multipliers;
5. Relax some of the active constraints (if needed);
6. Reduce penalization parameter ρ (if needed).

Outer Iteration : Regularization Method for the MPCC

Data: Let $z^0 = (x^0, s^0)$ be an initial point;
Choose a sequence of precision $\{\epsilon_k\}$ and a desired precision ϵ_∞;
Set $k = 0$;
Begin ;
repeat
    $(t, \bar{t})$:=Oracle(ϵ_k) ;
    Active-Set Algorithm : from the starting point z^k, use Algorithm Inner to compute z^{k+1} an approximate stationary point of $(Relax_{t,\bar{t}})$;
    Set $k \leftarrow k + 1$;
until x^{k+1} is M-stationary of the MPCC up to ϵ_∞;
return: f_{opt} the optimal value at the solution x_{opt} or a decision of infeasibility or unboundedness.

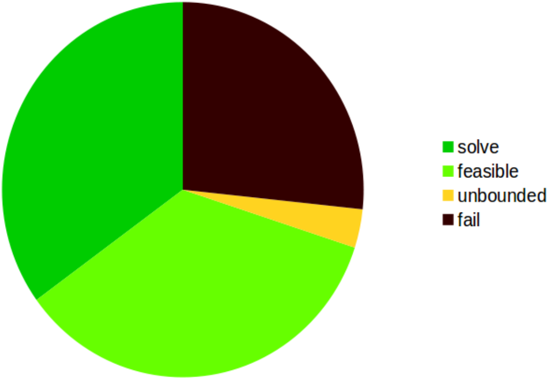
Julia

- Recent high level dynamic programming language (2012);
- Sophisticated compilation with performances close to C;
- Designed for high performance numerical analysis and computational science.

Already a lot of stuff available for optimization:

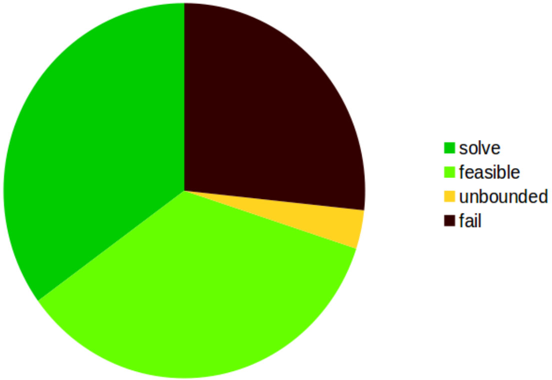
- JuMP (Modelling Language);
- MathProgBase (Interface between models and solvers);
- ...

MacMPEC test problems (60 pbs) with the butterfly relaxation for $\epsilon = 10^{-3}$



MacMPEC test problems

MacMPEC test problems (60 pbs) with the butterfly relaxation for $\epsilon = 10^{-3}$



Remark

Tests using a "naive straightforward" application of the algorithm.

Outline

Conclusions

We studied several regularization techniques for complementarity (and related problems). Our aim was to:

Conclusions

We studied several regularization techniques for complementarity (and related problems). Our aim was to:

1. study a general framework of methods
   - θ functions for ℓ_0 and CP
   - unified framework (UF) for MPCC

Conclusions

We studied several regularization techniques for complementarity (and related problems). Our aim was to:

1. study a general framework of methods
   - θ functions for ℓ_0 and CP
   - unified framework (UF) for MPCC
2. recover the best-known theoretical properties
   - M-stationarity for MPCC regularizations

Conclusions

We studied several regularization techniques for complementarity (and related problems). Our aim was to:

1. study a general framework of methods
   - θ functions for ℓ_0 and CP
   - unified framework (UF) for MPCC
2. recover the best-known theoretical properties
   - M-stationarity for MPCC regularizations
3. derive new theoretical results
   - sufficient condition for ℓ_0
   - error bounds for AVE
   - convergence, existence of stationary points for the MPCC

Conclusions

We studied several regularization techniques for complementarity (and related problems). Our aim was to:

1. study a general framework of methods
   - θ functions for ℓ_0 and CP
   - unified framework (UF) for MPCC
2. recover the best-known theoretical properties
   - M-stationarity for MPCC regularizations
3. derive new theoretical results
   - sufficient condition for ℓ_0
   - error bounds for AVE
   - convergence, existence of stationary points for the MPCC
4. overcome numerical difficulties
   - strong ϵ-stationary point
   - active-set penalization regularization strategy for the MPCC

Perspectives

1. Performance:
   - application to interior-point methods (Haddou, M., Migot, T., Omer, J. *A new direction in IPMs*, 2016)
   - penalization of the merit function (TAVE2 for AVE – Abdallah, L.,Haddou, M., Migot, T., *A DC Subadditive Approach for CP*, 2017)
   - solver in Julia for degenerate NLPs (including MPCCs)

Perspectives

1. Performance:
   - application to interior-point methods (Haddou, M., Migot, T., Omer, J. *A new direction in IPMs*, 2016)
   - penalization of the merit function (TAVE2 for AVE – Abdallah, L.,Haddou, M., Migot, T., *A DC Subadditive Approach for CP*, 2017)
   - solver in Julia for degenerate NLPs (including MPCCs)
2. Extension:
   - mathematical programs with vanishing constraints
   - optimization models with cardinality constraints

Perspectives

1. Performance:
   - application to interior-point methods (Haddou, M., Migot, T., Omer, J. *A new direction in IPMs*, 2016)
   - penalization of the merit function (TAVE2 for AVE – Abdallah, L.,Haddou, M., Migot, T., *A DC Subadditive Approach for CP*, 2017)
   - solver in Julia for degenerate NLPs (including MPCCs)
2. Extension:
   - mathematical programs with vanishing constraints
   - optimization models with cardinality constraints
3. Application:
   - bilevel programming

Thank you for your attention !

Outline

Using this technique, we tackle difficult problems such as:

1. Sparse Optimization,
   - Haddou, M., Migot, T., A Smoothing Method for Sparse Optimization, *Proceedings MCO*, 2015.

Outline

Using this technique, we tackle difficult problems such as:

1. Sparse Optimization,
   - Haddou, M., Migot, T., A Smoothing Method for Sparse Optimization, *Proceedings MCO*, 2015.
2. Complementarity Problems and related problems,
   - Abdallah, L., Haddou, M., Migot, T., Solving AVE using Complementarity and Smoothing Functions, *JCAM*, 2018.

Outline

Using this technique, we tackle difficult problems such as:

1. Sparse Optimization,
   - Haddou, M., Migot, T., A Smoothing Method for Sparse Optimization, *Proceedings MCO*, 2015.
2. Complementarity Problems and related problems,
   - Abdallah, L., Haddou, M., Migot, T., Solving AVE using Complementarity and Smoothing Functions, *JCAM*, 2018.
3. Mathematical Programs with Complementarity Constraints,
   - Dussault, J.-P., Haddou, M., Migot, T., The New Butterfly Relaxation for MPCC, *optimization-online.org*, 2016,
   - Dussault, J.-P., Haddou, M., Kadrani, A., Migot, T., How to Compute a Local Minimum of the MPCC, *optimization-online.org*, 2017.

Numerical results with random $n \times m$ matrix A

Successive Linearization Algorithm to solve the concave problem.

Compare a default sparse solution with 10% of non-zero components, $\#l_0$, the initial iterate solution of (P_1), $\#l_1$, and the solution by θ-algorithm with function θ^1, $\#\theta^1$.

n	m	$\#\ell_0 \geq \#\theta^1$	$\#\ell_1 = \#\ell_0$	$\#\theta^1 < \#\ell_1$
1000	800	100	100	0
1000	600	100	98	2
1000	400	50	1	99
750	600	100	100	0
750	450	100	98	2
750	300	54	0	100
500	400	100	100	0
500	300	100	94	6
500	200	63	0	100

θ Regularization of AVE: Numerics

- 100 problems for each size;
- A from a uniform distribution on $[-10, 10]$;
- x from a uniform distribution on $[-1, 1]$;
- $b = Ax - |x|$.

n	CMM	LPM	TAVE	TAVE2
32	9	7	0	0
64	8	13	3	2
128	10	13	8	4
256	11	11	8	4

Comparison

A sensitivity analysis on several values of the parameters, 7 values for t_0 and 3 for σ_t.

We take into account three criteria :

a) Feasibility of the last relaxed non-linear program: $\max(-g(x), |h(x)|, -\Phi(x)) \leq 10^{-7}$;

b) Feasibility of the complementarity constraint: $\min(G(x), H(x)) \leq \sqrt{10^{-7}}$;

(c) The complementarity between the Lagrange multipliers and the constraints of the last relaxed non-linear program.

MacMPEC Test Problems

MacMPEC is a collection of test problems from real-world applications available in AMPL.

Leyffer, Sven.
MacMPEC: AMPL collection of MPECs.
www.mcs.anl.gov/leyfier/MacMPEC, 2000.

Comparison

We run the simulation with three different solvers IPOPT, MINOS and SNOPT and present here the best.

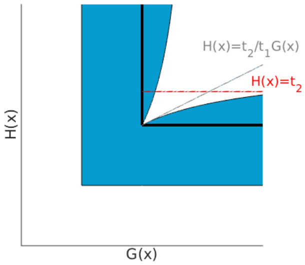
Results using SNOPT to solve the non-linear programs:

101 pb						
snopt	NL	SS	KS	$B_{(t=r)}$	$B_{(s=t,2t=r)}$	$B_{(t=r^{3/2})}$
best	92.1	94.1	94.1	**96.0**	93.1	**95.0**
average	92.1	90.4	90.3	**91.7**	89.4	**91.6**
worst	92.1	83.2	86.1	87.1	86.1	87.1

min: % worst set of parameter; **average**: average % of success; **max**: % best set of parameter

.

About the Butterfly Relaxation



Butterfly relaxation:
$\Phi(a, b; t) =$
$(a - t_1\theta_{t_2}(b))(b - t_1\theta_{t_2}(a))$.

Example

$$\min_{x \in \mathbb{R}^2} -x_1 \text{ s.t } x_1 \leq 1, \; 0 \leq x_1 \perp x_2 \geq 0.$$

- There are two stationary points: $(1, 0)^T$ S-stat. and $(0, 0)^T$ M-stat.;
- Relaxation KS and KDB: $x^k = (t_{2,k}, 2t_{2,k})^T \to (0, 0)^T$;
- There is no such sequence for the butterfly relaxation.

Existence of strong epsilon-stationary point

1. Let $x^* \in \mathcal{Z}$ be an M-stationary point;
2. $\epsilon > 0$ arbitrarily small;
3. Hypothesis on ψ (that encompass relaxations KS, Butterfly and approximation KDB);

Theorem (*Dussault-Haddou-Kadrani-Migot,17'*)

Then, there exists positive constants c, $\bar{t}^*$ with $\bar{t}^* > c\epsilon$ and a neighbourhood $U(x^*)$ of $(x^*, G(x^*), H(x^*))^T$ such that for all $t \in (0, t^*)$ and $\bar{t} \in (0, \bar{t}^*)$ there exists $(x, s)^T \in U(x^*)$, which is strong ϵ-stationary point of the relaxation with slack variables.

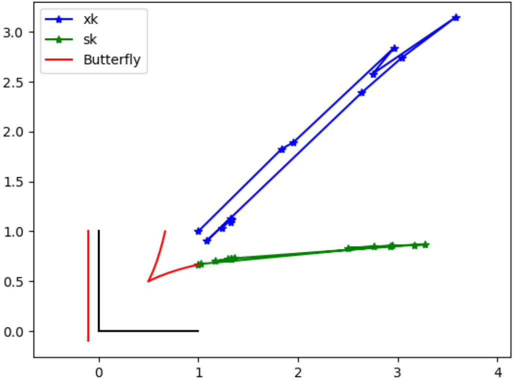
Counter-example without slack variables.

Example: Step I

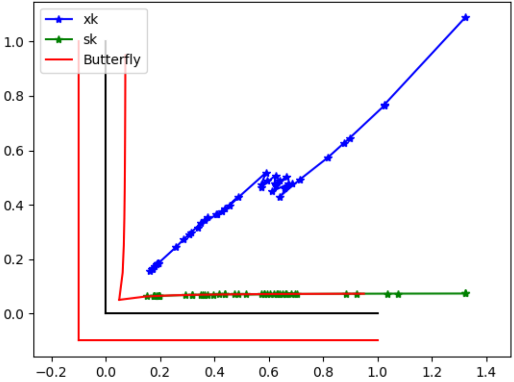
$$\min_{x\in\mathbb{R}^2} x_1^2 + x_2^2 - x_3 \text{ s.t } s_G = x_1,\ s_H = x_2, 0 \leq s_G \perp s_H \geq 0,$$
$$-4x_1 + x_3 \leq 0,\ -4x_2 + x_3 \leq 0.$$

Example: Step I

$$\min_{x\in\mathbb{R}^2} x_1^2 + x_2^2 - x_3 \text{ s.t } s_G = x_1,\ s_H = x_2, 0 \le s_G \perp s_H \ge 0,$$
$$-4x_1 + x_3 \le 0,\ -4x_2 + x_3 \le 0.$$



$$\min_{x \in \mathbb{R}^2} x_1^2 + x_2^2 - x_3 \text{ s.t } s_G = x_1,\ s_H = x_2, 0 \leq s_G \perp s_H \geq 0,$$
$$-4x_1 + x_3 \leq 0,\ -4x_2 + x_3 \leq 0.$$



Example: Step III

$$\min_{x\in\mathbb{R}^2} x_1^2 + x_2^2 - x_3 \text{ s.t } s_G = x_1,\ s_H = x_2, 0 \leq s_G \perp s_H \geq 0,$$
$$-4x_1 + x_3 \leq 0,\ -4x_2 + x_3 \leq 0.$$