Formulations and solution algorithms for Complementarity Problems (or seven ways to skin a cat)

Michael C. Ferris

University of Wisconsin, Madison

Funded by AFOSR and NSF

Zinal, January 19, 2017

AVI over polyhedral convex set

An affine function

$$F : \mathbb{R}^n \to \mathbb{R}^n,\ F(z) = Mz + q,\ M \in \mathbb{R}^{n \times n},\ q \in \mathbb{R}^n$$

A polyhedral convex set

$$\mathcal{C} = \{z \in \mathbb{R}^n \mid Az(\geq, =, \leq)a,\ l \leq z \leq u\},\ A \in \mathbb{R}^{m \times n}$$

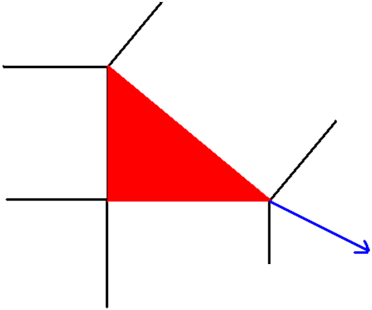
Find a point $z^* \in \mathcal{C}$ satisfying

$$\begin{aligned} \langle F(z^*), y - z^* \rangle \quad & \geq 0, \quad \forall y \in \mathcal{C} \\ (\Leftrightarrow)\ \langle -F(z^*), y - z^* \rangle \quad & \leq 0, \quad \forall y \in \mathcal{C} \\ (\Leftrightarrow)\ -F(z^*) \quad & \in N_{\mathcal{C}}(z^*) \end{aligned}$$

where

$$N_{\mathcal{C}}(z^*) \quad = \{v \mid \langle v, y - z^* \rangle \leq 0, \forall y \in \mathcal{C}\}$$

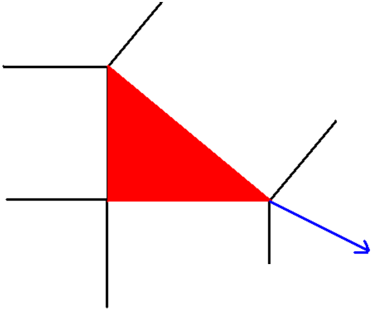
Normal map for polyhedral C



projection: $\pi_C(x)$

$$x - \pi_C(x) \in N_C(\pi_C(x))$$

Normal map for polyhedral C



projection: $\pi_C(x)$

$$x - \pi_C(x) \in N_C(\pi_C(x))$$

If $-M\pi_C(x) - q = x - \pi_C(x)$ then
So $z = \pi_C(x)$ solves

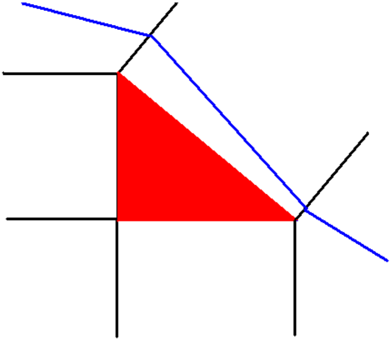
$$0 \in M\pi_C(x) + q + N_C(\pi_C(x))$$

if and only if we can find x, a zero of the normal map:

$$0 = M\pi_C(x) + q + x - \pi_C(x)$$

The PATHAVI algorithm

- Start in cell that has interior (face is an extreme point, so normal cone has interior - primary ray)
- Move towards a zero of affine map in cell
- Update direction when hit boundary (pivot)
- Solves, or determines infeasible if M is copositive-plus on rec(C)
- Solves 2-person bimatrix games, 3-person games too, but these are nonlinear



But algorithm has exponential complexity (von Stengel et al)

Theorem

Suppose $\mathcal{C}$ is a polyhedral convex set and M is an L–matrix with respect to $\mathrm{rec}\mathcal{C}$ which is invertible on the lineality space of $\mathcal{C}$. Then exactly one of the following occurs:

- *PATHAVI solves (AVI)*
- *the following system has no solution*

$$Mz + q \in (\mathrm{rec}\mathcal{C})^D, \qquad z \in \mathcal{C}. \tag{1}$$

Corollary

If M is copositive–plus with respect to $\mathrm{rec}\mathcal{C}$, then exactly one of the following occurs:

- *PATHAVI solves (AVI)*
- *(1) has no solution*

Note also that if $\mathcal{C}$ is compact, then any matrix M is an L–matrix with respect to $\mathrm{rec}\mathcal{C}$. So always solved.

Experimental results: AVI vs MCP

PATH is a solver for MCP (mixed complementarity problem).

- Run PathAVI over AVI formulation.
- Run PATH over AVI in MCP form (poorer theory as $\text{rec}\mathcal{C}$ larger).
- Data generation
  - M is an $n \times n$ symmetric positive definite/indefinite matrix.
  - A has m randomly generated bounded inequality constraints.

(m, n)	PathAVI		PATH		% negative eigenvalues
	status	# iterations	status	# iterations	
(180,60)	S	55	S	72	0
(180,60)	S	45	S	306	20
(180,60)	S	2	F	9616	60
(180,60)	S	1	F	10981	80
(360,120)	S	124	S	267	0
(360,120)	S	55	S	1095	20
(360,120)	S	2	F	10020	60
(360,120)	S	1	F	7988	80

Extension to Nonlinear Model

- So now we can solve AVI, what happens when F is nonlinear
- Embed AVI solver in a Newton Method - each Newton step solves an AVI
- Nonlinear equations $F(x) = 0$
- Newton's Method

$$F(x^k) + \nabla F(x^k) d^k = 0$$
$$x^{k+1} = x^k + d^k$$

- Damp using Armijo linesearch on $\frac{1}{2}\|F(x)\|_2^2$
- Descent direction - gradient of merit function
- Properties
  - Well defined
  - Global and local-fast convergence

Nonsmooth Newton Method

Given x^k

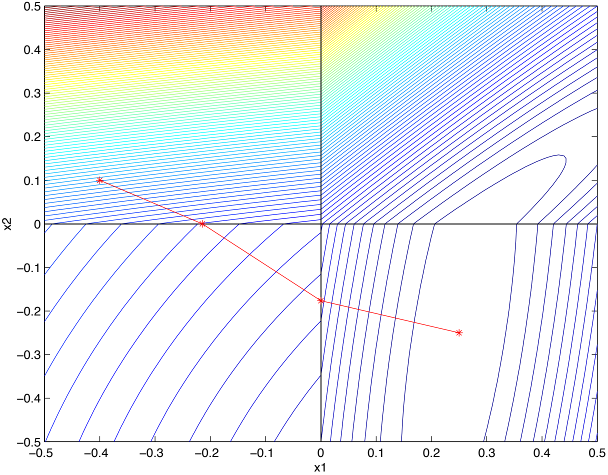
$$\text{solve: } 0 \in F(x^k) + \nabla F(x^k)(x - x^k) + N_C(x)$$
$$d_k = x^* - x^k, \ x^* \text{ from above}$$
$$x^{k+1} = x^k + \alpha d^k$$

- Equivalent piecewise smooth equation $F_+(x) = 0$

$$F_+(x) \equiv F(\pi_C(x)) + x - \pi_C(x)$$

  (when $C = \mathbb{R}^n_+$ then $\pi_C(x) = max(x, 0)$ is easy to compute)
- Nonsmooth Newton Method
  - Iteratively solve piecewise linear system of equations, via pivoting
  - Damp using Armijo search on $\frac{1}{2}\|F_+(x)\|_2^2$
- Properties
  - Global and local-fast convergence
  - Merit function *not* differentiable

Piecewise Linear Example



Fischer-Burmeister Function

$$\phi(a,b) := \sqrt{a^2+b^2} - a - b$$

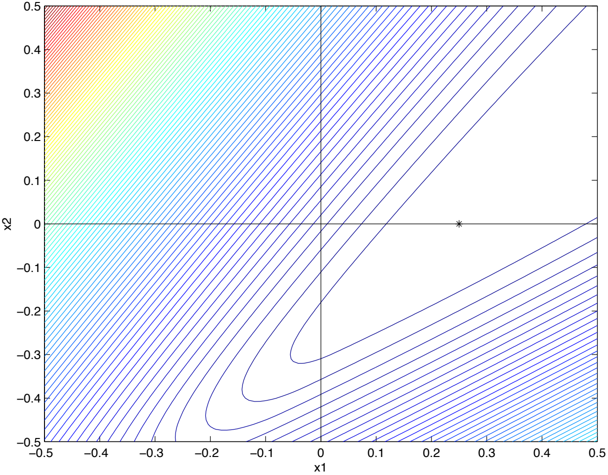
$$\phi(a,b) = 0 \iff 0 \le a \perp b \ge 0$$

- $\Phi(x)$ defined componentwise

$$\Phi_i(x) \equiv \sqrt{(x_i)^2 + (F_i(x))^2} - x_i - F_i(x)$$

- $\Phi(x) = 0$ if and only if x solves NCP(F)
- Not continuously differentiable - semismooth
- Natural merit function ($\frac{1}{2}\|\Phi(x)\|_2^2$) is differentiable

Fischer-Burmeister Example



Review

- Nonlinear Complementarity Problem
- Piecewise smooth system of equations
  - Use nonsmooth Newton Method
  - Solve linear complementarity problem per iteration
  - Merit function not differentiable
- Fischer-Burmeister
  - Differentiable merit function
- Combine to obtain new algorithm
  - Well defined
  - Global and local-fast convergence

Feasible Descent Framework

- Calculate direction using a local method
  - Generates feasible iterates
  - Local fast convergence
  - Used nonsmooth Newton Method
- Accept direction if descent for $\frac{1}{2}\|\Phi(x)\|^2$
- Otherwise use projected gradient step

Theorem

Let $\{x^k\} \subseteq \Re^n$ be a sequence generated by the algorithm that has an accumulation point x^* which is a strongly regular solution of the NCP. Then the entire sequence $\{x^k\}$ converges to this point, and the rate of convergence is Q-superlinear.

- Method is well defined
- Accumulation points are stationary points
- Locally projected gradient steps not used

Computational Details

- Preprocessing to simplify without changing underlying problem
- Crashing method to quickly identify basis
- Nonmonotone search with watchdog
- Perturbation scheme for rank deficiency
- Stable interpolating pathsearch
- Restart strategy
- Projected gradient searches

Nonlinear Complementarity Problems

- Given $F : \Re^n \to \Re^n$
- Find $x \in \Re^n$ such that

$$0 \leq F(x) \qquad x \geq 0$$

$$x^T F(x) = 0$$

- Compactly written

$$0 \leq F(x) \quad \perp \quad x \geq 0$$

- Equivalent to nonsmooth equation (min-map):

$$\min(x, F(x)) = 0$$

Nonsmooth alternatives

The normal map is one nonsmooth equation reformulation of the nonlinear complementarity problem.

We have just seen two alternatives

1. Fischer-Burmeister $\Phi(x) = 0$
2. Min-map $\min(x, F(x)) = 0$

Alternative methods generate generalized derivatives of these nonsmooth functions and use within nonsmooth Newton methods

- Approaches are relatively simple to implement and work well in many (well defined) cases
- Fundamental difference is nonsmoothness is outside F
- PATH tends to perform better (due to the heuristic extensions) on harder/messier problems

Smoothing: The Fischer Function [Burmeister]

- For NCP (with $\mu > 0$):

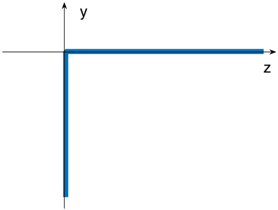
$$0 = \phi_\mu(x_i, F_i(x)), \quad i = 1, 2, \ldots, n$$

where

$$\phi_\mu(a, b) := \sqrt{a^2 + b^2 + \mu} - a - b$$

- Gives rise to semismooth algorithms
- Need to drive μ to 0, no longer nonsmooth
- Available within NLPEC

Complementarity Problems via Graphs



- $\mathcal{T} = \mathcal{N}_{\mathbb{R}_+} = (\mathbb{R}_+ \times \{0\}) \bigcup (\{0\} \times \mathbb{R}_-)$
- $\mathcal{T}$ is "monotone"

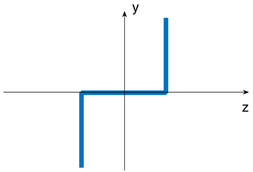
$$-y \in \mathcal{T}(z) \iff (z, -y) \in \mathcal{T} \iff 0 \le y \perp z \ge 0$$

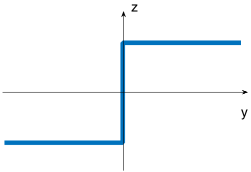
By approximating (smoothing) graph can generate interior point algorithms for example $yz = \epsilon, y, z > 0$

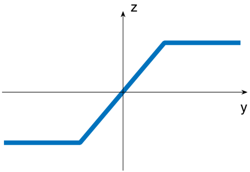
$$0 \in F(z) + \mathcal{N}_{\mathbb{R}^n_+}(z) \iff (z, -F(z)) \in \mathcal{T}^n \iff 0 \le F(z) \perp z \ge 0$$

Operators and Graphs ($\mathcal{C} = [-1, 1], \mathcal{T} = \mathcal{N}_\mathcal{C}$)

$$z_i = -1, -F_i(z) \leq 0 \text{ or } z_i \in (-1, 1), -F_i(z) = 0 \text{ or } z_i = 1, -F_i(z) \geq 0$$







$\mathcal{T}(z)$ $\quad$ $\mathcal{T}^{-1}(y)$ $\quad$ $(\mathcal{I} + \mathcal{T})^{-1}(y) = P_\mathcal{T}(y)$

$P_\mathcal{T}(y)$ is the projection of y onto $[-1, 1]$

Generalized Equations

- Suppose $\mathcal{T}$ is a maximal monotone operator

$$0 \in F(z) + \mathcal{T}(z) \qquad (GE)$$

- Define $P_{\mathcal{T}} = (\mathcal{I} + \mathcal{T})^{-1}$
- If $\mathcal{T}$ is polyhedral (graph of $\mathcal{T}$ is a finite union of convex polyhedral sets) then $P_{\mathcal{T}}$ is piecewise affine (continous, single-valued, non-expansive)

$$\begin{aligned} 0 \in F(z) + \mathcal{T}(z) \quad &\iff \quad z \in F(z) + \mathcal{I}(z) + \mathcal{T}(z) \\ &\iff \quad z - F(z) \in (\mathcal{I} + \mathcal{T})(z) \iff P_{\mathcal{T}}(z - F(z)) = z \end{aligned}$$

Use in fixed point iterations (cf projected gradient methods): this is in fact just the min-map!

Normal Map

- Suppose $\mathcal{T}$ is a maximal monotone operator

$$0 \in F(z) + \mathcal{T}(z) \qquad (GE)$$

- Define $P_{\mathcal{T}} = (I + \mathcal{T})^{-1}$ (continuous, single-valued, non-expansive)

$$\begin{aligned} 0 \in F(z) + \mathcal{T}(z) &\iff z \in F(z) + \mathcal{I}(z) + \mathcal{T}(z) \\ &\iff z - F(z) = x \text{ and } x \in (\mathcal{I} + \mathcal{T})(z) \\ &\iff z - F(z) = x \text{ and } P_{\mathcal{T}}(x) = z \\ &\iff P_{\mathcal{T}}(x) - F(P_{\mathcal{T}}(x)) = x \\ &\iff 0 = F(P_{\mathcal{T}}(x)) + x - P_{\mathcal{T}}(x) \end{aligned}$$

This is the so-called Normal Map Equation

Splitting Methods

- Suppose $\mathcal{T}$ is a maximal monotone operator

$$0 \in F(z) + \mathcal{T}(z) \qquad (GE)$$

- Can devise Newton methods (e.g. SQP) that treat F via calculus and $\mathcal{T}$ via convex analysis
- Alternatively, can split $F(z) = A(z) + B(z)$ (and possibly $\mathcal{T}$ also) so we solve solve (GE) by solving a sequence of problems involving just

$$\mathcal{T}_1(z) = A(z) \text{ and } \mathcal{T}_2(z) = B(z) + \mathcal{T}(z)$$

  where each of these is "simpler"
- Forward-Backward splitting (or ADMM):

$$z^{k+1} = (I + c_k T_2)^{-1} (I - c_k T_1) \left(z^k\right),$$

The problem

- MOPEC: x^*, y:

$$x_i^* \text{ solves } \min_{x_i \in K_i(x_{-i}^*, y)} \theta(x_i, x_{-i}^*, y), \forall i$$

$$y \text{ solves } VI(F(x^*, \cdot), C)$$

$$\begin{bmatrix} \mathcal{A}_1 & A_{1,2} & \cdots & A_{1,p} & E_1 \\ A_{2,1} & \mathcal{A}_2 & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & A_{p-1,p} & E_{p-1} \\ A_{p,1} & \cdots & A_{p,p-1} & \mathcal{A}_p & E_p \\ F_1 & \cdots & F_{p-1} & F_p & D \end{bmatrix}$$

Strongly Convex Nash Equilibria

$$\min_{x_1 \geq 0} \ \frac{1}{2}x_1^2 - \theta x_1 x_2 - 4x_1 \text{ s.t. } x_1 + x_2 \geq 1$$

$$\min_{x_2 \geq 0} \ \frac{1}{2}x_2^2 - x_1 x_2 - 3x_2$$

- No solution for $\theta \geq 1$:

$$x_1(x_2) = (\theta x_2 + 4)_+, \quad x_2(x_1) = (x_1 + 3)_+$$

- Solution $-\frac{4}{3} \leq \theta < 1$: $x_1 = \frac{4+3\theta}{1-\theta}$, $x_2 = x_1 + 3$
- Solution $\theta \leq -\frac{4}{3}$: $x_1 = 0$, $x_1 = 3$
- Jacobi works provided $\theta < 1$, but theory fails

The Issues

This is not the optimality conditions of a single optimization problem:

$$0 \leq \left[\begin{array}{cc|c} 1 & 1 & -\theta \\ 1 & 0 & 1 \\ \hline -1 & & 1 \end{array}\right] \begin{bmatrix} x_1 \\ -p_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} 4 \\ 1 \\ 3 \end{bmatrix} \quad \perp \quad \begin{bmatrix} x_1 \\ -p_1 \\ x_2 \end{bmatrix} \geq 0$$

- The matrix $\mathcal{A}$ in general is never diagonally dominant except in trivial cases
- Iterations based on succesive inversion of local blocks (or successive optimization of local strategies) can converge.
- We establish sufficient conditions which guarantee convergence of block Jacobi and block Gauss-Seidel iterations for such matrices.

Economic Application

- Model is a partial equilibrium, geographic exchange model.
- Goods are distinguished by region of origin.
- There is one unit of region r goods.
- These goods may be consumed in region r or they may be exported.
- Each region solves:

$$\min_{X,T_r} f_r(X,T) \text{ s.t. } F(X,T) = 0,\ T_j = \bar{T}_j, j \neq r$$

  where $f_r(X,T)$ is a quadratic form and $F(X,T)$ is linear and defines X uniquely as a function of T.
- $F(X,T)$ defines an equilibrium; here it is simply a set of equations, not a complementarity problem

Results

Gauss-Seidel residuals

Iteration	deviation
1	3.14930
2	0.90970
3	0.14224
4	0.02285
5	0.00373
6	0.00061
7	0.00010
8	0.00002
9	0.00000

Tariff revenue

region	SysOpt	MOPEC
1	0.117	0.012
2	0.517	0.407
3	0.496	0.214
4	0.517	0.407
5	0.117	0.012

- Note that competitive solution produces much less revenue than system optimal solution
- Model has non-convex objective, but each subproblem is solved globally (lindoglobal)

MIP formulations for Complementarity

Set $y_i = F_i(x)$, then additionally

$$y_i \geq 0, x_i \geq 0, x_i y_i = 0$$

If we know upper bounds on x_i and y_i we can model as:

$$(x_i, y_i) \in SOS1$$

or introduce binary variable z_i and

$$x_i \leq Mz_i, y_i \leq M(1 - z_i)$$

(or use indicator variables to turn on "fixing" constraints). Works if bounds are good and problem size is not too large. Issues with bounds on multipliers not being evident.

MPEC approaches

- Can use nonlinear programming approaches (e.g. NLPEC (see previous lecture))
- Knitro can process MPCC's and uses penalization for complementarity
- Implicit approach: generate $y(x)$ where y solves the parametric (in x) complementarity problem, then solve

$$\min f(x, y(x))$$

using a bundle trust region method for example. Difficult to deal with side constraints.

Conclusions

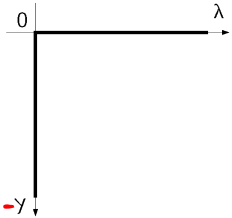
- Many formulations and algorithms for complementarity problems
- PATH algorithm is widely used, available in GAMS, AMPL, AIMMS, JUMP, Matlab, API-format
- Need for more theoretic and algorithmic enhancements in large scale and structured cases
- Need to find all solutions of complementarity problems, or to solve MPEC/MPCC to global optimality

Complementarity Systems (DVI)

$$\frac{dx}{dt}(t) = f(x(t), z(t))$$

$$y(t) = h(x(t), z(t))$$

$$0 \leq y(t) \perp z(t) \geq 0$$

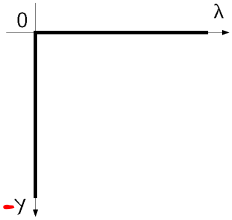


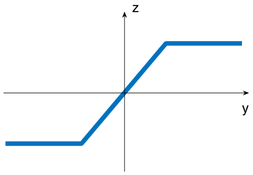
Complementarity Systems (DVI)

$$\frac{dx}{dt}(t) = f(x(t), z(t))$$

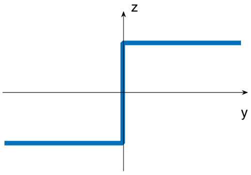
$$y(t) = h(x(t), z(t))$$

$$0 \leq y(t) \perp z(t) \geq 0$$

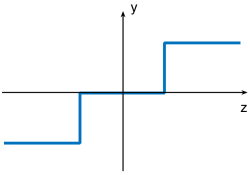




saturation



relay



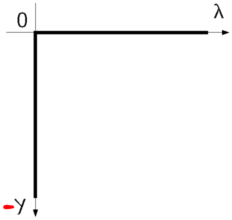
relay with dead zone

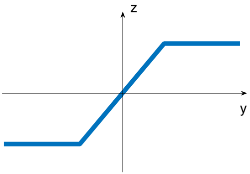
Complementarity Systems (DVI)

$$\frac{dx}{dt}(t) = f(x(t), z(t))$$

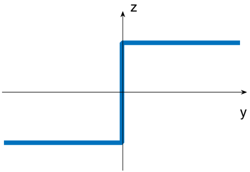
$$y(t) = h(x(t), z(t))$$

$$(z(t), -y(t)) \in \mathcal{T}$$

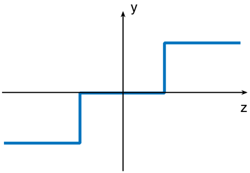




saturation



relay



relay with dead zone

Separable Structure

- Partition variables into (x, y)
- Identify separable structure

$$0 \in \begin{bmatrix} F(x) \\ G(x,y) \end{bmatrix} + \begin{bmatrix} N_{\Re^n_+}(x) \\ N_{\Re^n_+}(y) \end{bmatrix}$$

- Reductions possible if either
  1. $0 \in F(x) + N_{\Re^n_+}(x)$ has a unique solution
  2. $0 \in G(x,y) + N_{\Re^n_+}(y)$ has solution for all x
- Theory provides appropriate conditions
- Solve F and G sequentially