Optimality conditions and complementarity, variational inequalities, operators and graphs

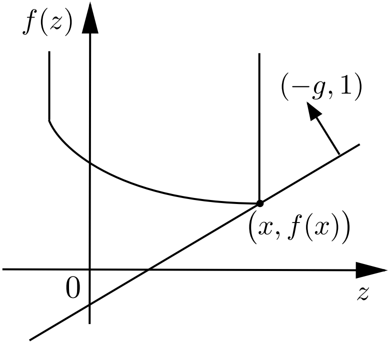
Michael C. Ferris

University of Wisconsin, Madison

Funded by DOE-MACS Grant with Argonne National Laboratory

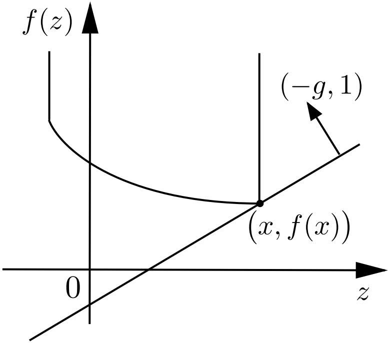
Zinal, January 16, 2017

Convex subdifferentials



- Assume f is convex, then $f(z) \geq f(x) + \nabla f(x)^T(z - x)$ (linearization is below the function)
- Incorporate constraints by allowing f to take on $+\infty$ if constraint is violated $f : \mathbb{R}^n \mapsto (-\infty, +\infty]$
- $\partial f(x) = \{g : f(z) \geq f(x) + g^T(z - x), \forall z\}$, the subdifferential of f at x

Convex subdifferentials



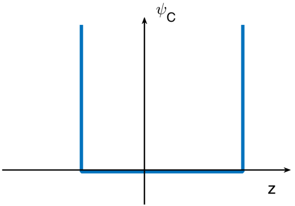
- Assume f is convex, then $f(z) \geq f(x) + \nabla f(x)^T(z - x)$ (linearization is below the function)
- Incorporate constraints by allowing f to take on $+\infty$ if constraint is violated $f : \mathbb{R}^n \mapsto (-\infty, +\infty]$
- $\partial f(x) = \{g : f(z) \geq f(x) + g^T(z - x), \forall z\}$, the subdifferential of f at x

- If f is differentiable and convex, then $\partial f(x) = \{\nabla f(x)\}$
- e.g. $f(z) = \frac{1}{2}z^T Qz + p^T z$, then $\partial f(x) = \{Qx + p\}$
- x^* solves $\min f(x)$ if and only if $0 \in \partial f(x^*)$

Indicator functions and normal cones

$$\psi_{\mathcal{C}}(z) = \begin{cases} 0 & \text{if } z \in \mathcal{C} \\ \infty & \text{else} \end{cases}$$

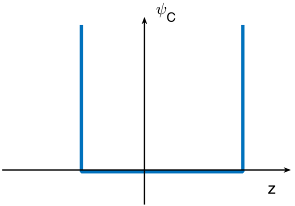
$\psi_{\mathcal{C}}$ is a convex function when $\mathcal{C}$ is a convex set

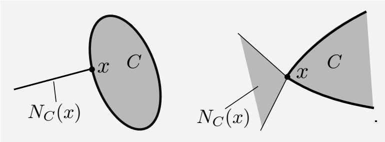


Indicator functions and normal cones

$$\psi_{\mathcal{C}}(z) = \begin{cases} 0 & \text{if } z \in \mathcal{C} \\ \infty & \text{else} \end{cases}$$

$\psi_{\mathcal{C}}$ is a convex function when $\mathcal{C}$ is a convex set





If $x \in \mathcal{C}$, then

$$g \in \partial\psi_{\mathcal{C}}(x)$$
$$\iff \psi_{\mathcal{C}}(z) \geq \psi_{\mathcal{C}}(x) + g^T(z-x), \ \forall z$$
$$\iff 0 \geq g^T(z-x), \ \forall z \in \mathcal{C}$$

Normal cone to $\mathcal{C}$ at x,

$$N_{\mathcal{C}}(x) := \partial\psi_{\mathcal{C}}(x) = \begin{cases} \{g : g^T(z-x) \leq 0, \forall z \in \mathcal{C}\} & \text{if } x \in \mathcal{C} \\ \emptyset & \text{if } x \notin \mathcal{C} \end{cases}$$

Some calculus

- $f_i : \mathbb{R}^n \mapsto (-\infty, \infty]$, $i = 1, \ldots, m$, proper, convex functions

$$F = f_1 + \cdots + f_m$$

assume $\bigcap_{i=1}^{m} \operatorname{rint}(\operatorname{dom}(f_i)) \neq \emptyset$ then (as sets)

$$\partial F(x) = \partial f_1(x) + \cdots + \partial f_m(x),\ \forall x$$

- $\mathcal{C} = \bigcap_{i=1}^{m} \mathcal{C}_i$, then $\psi_{\mathcal{C}} = \psi_{\mathcal{C}_i} + \cdots + \psi_{\mathcal{C}_m}$, so $N_{\mathcal{C}} = N_{\mathcal{C}_i} + \cdots + N_{\mathcal{C}_m}$

$$x^* \text{ solves } \min_{x \in \mathcal{C}} f(x) \iff x^* \text{ solves } \min_x (f + \psi_{\mathcal{C}})(x)$$

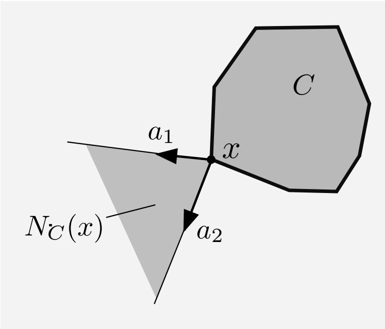
$$\iff 0 \in \partial(f + \psi_{\mathcal{C}})(x^*) \iff 0 \in \nabla f(x^*) + N_{\mathcal{C}}(x^*)$$

Special cases and examples

- Normal cone is a cone
- $x \in \text{int}(\mathcal{C})$, then $N_{\mathcal{C}}(x) = \{0\}$
- $\mathcal{C} = \mathbb{R}^n$, then $N_{\mathcal{C}}(x) = \{0\}$, $\forall x \in \mathcal{C}$

Special cases and examples

- Normal cone is a cone
- $x \in \mathrm{int}(\mathcal{C})$, then $N_{\mathcal{C}}(x) = \{0\}$
- $\mathcal{C} = \mathbb{R}^n$, then $N_{\mathcal{C}}(x) = \{0\}, \forall x \in \mathcal{C}$



- $\mathcal{C} = \left\{z : a_i^T z \leq b_i, i = 1, \ldots, m\right\}$ polyhedral
- $N_{\mathcal{C}}(x) = \left\{ \sum_{i=1}^{m} \lambda_i a_i : 0 \leq b_i - a_i^T x \perp \lambda_i \geq 0 \right\}$
- $\perp$ makes product of items around it 0, i.e.

$$(b_i - a_i^T x)\lambda_i = 0, \; i = 1, \ldots, m$$

Combining: KKT conditions

- Example: convex optimization first-order optimality condition:

$$
\begin{aligned}
x^* \text{ solves } \min_{x \in \mathcal{C}} f(x) \iff & 0 \in \nabla f(x^*) + N_{\mathcal{C}}(x^*) \\
\iff & 0 = \nabla f(x^*) + y,\ y \in N_{\mathcal{C}}(x^*) \\
\iff & 0 = \nabla f(x^*) + y,\ y = A^T \lambda, \\
& 0 \le b - Ax^* \perp \lambda \ge 0 \\
\iff & 0 = \nabla f(x^*) + A^T \lambda, \\
& 0 \le b - Ax^* \perp \lambda \ge 0
\end{aligned}
$$

- More generally, if $\mathcal{C} = \{z : g(z) \le 0\}$, g convex, (with CQ)

$$
\begin{aligned}
x^* \text{ solves } \min_{x \in \mathcal{C}} f(x) \iff & 0 \in \nabla f(x^*) + N_{\mathcal{C}}(x^*) \\
\iff & 0 = \nabla f(x^*) + \nabla g(x^*) \lambda, \\
& 0 \le -g(x^*) \perp \lambda \ge 0
\end{aligned}
$$

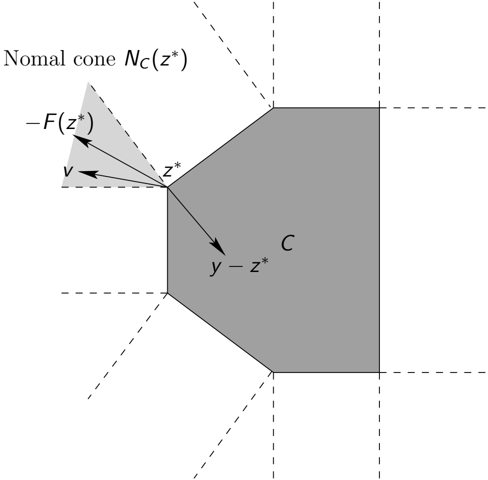
Variational Inequality (replace $\nabla f(z)$ with $F(z)$)

- $F : \mathbb{R}^n \to \mathbb{R}^n$
- Ideally: $\mathcal{C} \subseteq \mathbb{R}^n$ – constraint set; Often: $\mathcal{C} \subseteq \mathbb{R}^n$ – simple bounds

$$VI(F, \mathcal{C}) : \quad 0 \in F(z) + N_{\mathcal{C}}(z)$$

- VI generalizes many problem classes
- Nonlinear Equations: $F(z) = 0$ set $\mathcal{C} \equiv \mathbb{R}^n$
- Convex optimization: $F(z) = \nabla f(z)$
- For NCP: $0 \leq F(z) \perp z \geq 0$, set $\mathcal{C} \equiv \mathbb{R}^n_+$
- For MCP (rectangular VI), set $\mathcal{C} \equiv [l, u]^n$.
- For LP, set $F(z) \equiv \nabla f(z) = p$ and $\mathcal{C} = \{z : Az = a, Hz \leq h\}$.

VI: $0 \in F(z) + \mathcal{N}_{\mathcal{C}}(z)$



Many applications where F is not the derivative of some f

Other applications of complementarity

- Economics: Walrasian equilibrium (supply equals demand), taxes and tariffs, computable general equilibria, option pricing (electricity market), airline overbooking
- Transportation: Wardropian equilibrium (shortest paths), selfish routing, dynamic traffic assignment
- Applied mathematics: Free boundary problems
- Engineering: Optimal control (ELQP)
- Mechanics: Structure design, contact problems (with friction)
- Geology: Earthquake propogation

Tradeoff accuracy and simple structure

Many models from statistics: e.g. regression:

$$\min_x \|Ax - y\|^2$$

Additional structure: Compressed sensing: sparse signal to account for y

$$\min_x \|Ax - y\|_2^2 \text{ s.t. } \|x\|_0 \le c$$

Regularized regression:

$$\min_x \|Ax - y\|_2^2 + \alpha \|x\|_1$$

Tradeoff accuracy and simple structure

Many models from statistics: e.g. regression:

$$\min_x \|Ax - y\|^2$$

Additional structure: Compressed sensing: sparse signal to account for y

$$\min_x \|Ax - y\|_2^2 \text{ s.t. } \|x\|_0 \leq c$$

Regularized regression:

$$\min_x \|Ax - y\|_2^2 + \alpha \|x\|_1$$

Machine learning: SVM for classification

$$\min_{w,\xi,\gamma} \sum_i \xi_i + \frac{\alpha}{2} \|w\|^2 \text{ s.t. } D(Aw - \gamma 1) \geq 1 - \xi$$

General model:

$$\min_{x \in X} E(x) + \alpha S(x)$$

X are constraints, E measures “error” and S penalizes bad structure

Image denoising (Wright)

Rudin-Osher-Fatemi (ROF) model (ℓ_2−TV). Given a domain $\Omega \subset \mathbb{R}^2$ and an observed image $f : \Omega \mapsto \mathbb{R}$, seek a restored image $u : \Omega \mapsto \mathbb{R}$ that preserves edges while removing noise. The regularized image u can typically be stored more economically. Seek to "minimize" both

- $\|u - f\|_2$ and
- the total-variation (TV) norm $\int_\Omega |\nabla u| \, dx$

Use constrained formulations, or a weighting of the two objectives:

$$\min_u P(u) := \|u - f\|_2^2 + \alpha \int_\Omega |\nabla u| \, dx$$

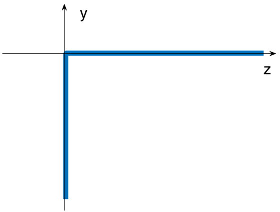
The minimizing u tends to have regions in which u is constant ($\nabla u = 0$). More "cartoon-like" when α is large.

Original, noisy, denoised (tol $= 10^{-2}$, 10^{-4})

Conclusions

- Convexity separates easy optimization problems from hard ones
- Modern convex analysis extends linear programming to richer but still tractable settings
- Modern optimization within applications requires multiple model formats, computational tools and sophisticated solvers
- Variational inequalities and set valued analysis important tools for big data problems
- Modeling, optimization, statistics and computation embedded within the application domain is critical
- Many new settings available for deployment; need for more theoretic and algorithmic enhancements

Complementarity Problems via Graphs



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

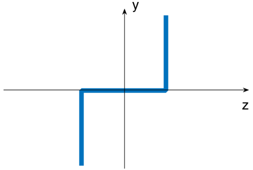
$$-y \in \mathcal{T}(z) \iff (z, -y) \in \mathcal{T} \iff 0 \leq y \perp z \geq 0$$

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

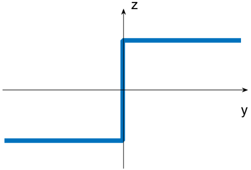
$$0 \in F(z) + \mathcal{N}_{\mathbb{R}^n_+}(z) \iff (z, -F(z)) \in \mathcal{T}^n \iff 0 \leq F(z) \perp z \geq 0$$

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

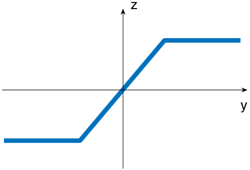
$$z_i = -1, -F_i(z) \le 0 \text{ or } z_i \in (-1, 1), -F_i(z) = 0 \text{ or } z_i = 1, -F_i(z) \ge 0$$



$\mathcal{T}(z)$



$\mathcal{T}^{-1}(y)$



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

$P_{\mathcal{T}}(y)$ 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)

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):

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

Normal Map

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

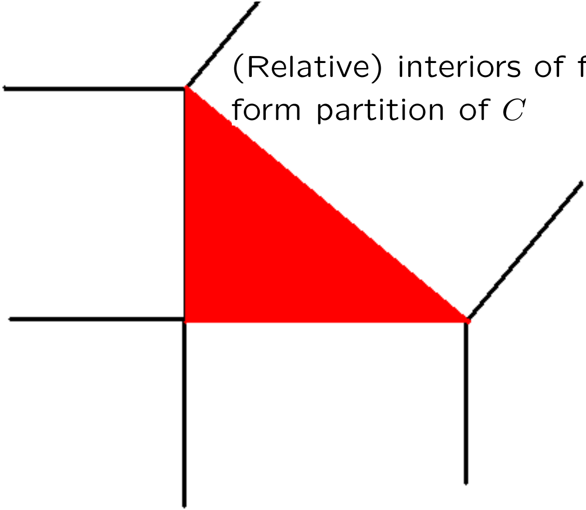
$$0 \in F(z) + \mathcal{T}(z) \qquad (GE)$$

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

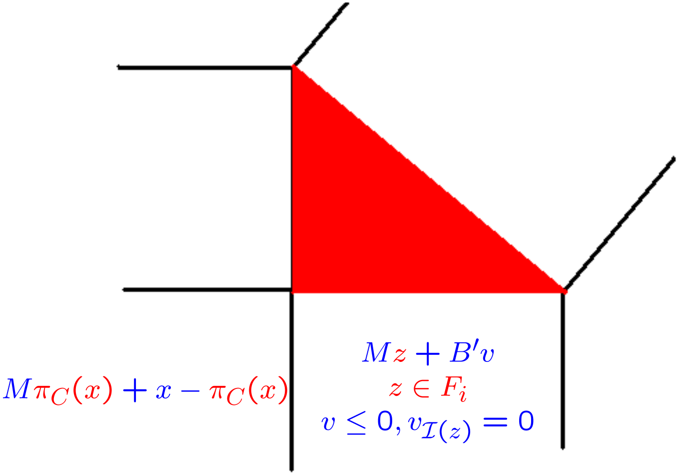
$$\begin{aligned}
0 \in F(z) + \mathcal{T}(z) \quad & \iff \quad z \in F(z) + \mathcal{I}(z) + \mathcal{T}(z) \\
& \iff \quad z - F(z) = x \text{ and } x \in (\mathcal{I} + \mathcal{T})(z) \\
& \iff \quad z - F(z) = x \text{ and } P_{\mathcal{T}}(x) = z \\
& \iff \quad P_{\mathcal{T}}(x) - F(P_{\mathcal{T}}(x))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

This is the so-called Normal Map Equation

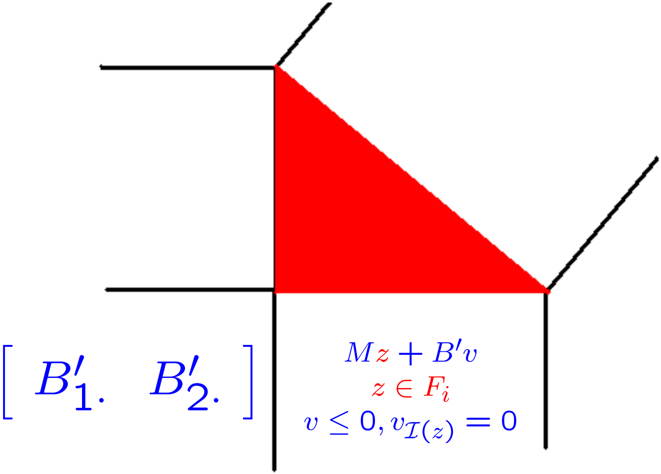
Normal manifold = $\{F_i + N_{F_i}\}$



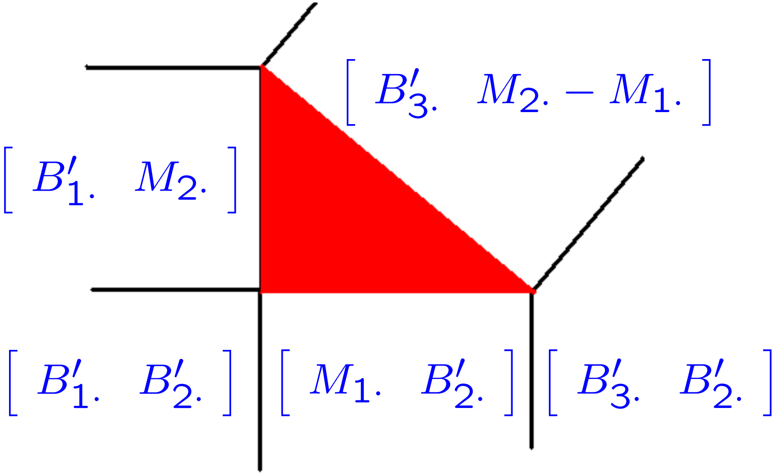
$C = \{z | Bz \geq b\}, N_C(z) = \{B'v | v \leq 0, v_{\mathcal{I}(z)} = 0\}$



$$C = \{z | Bz \geq b\}, N_C(z) = \{B'v | v \leq 0, v_{\mathcal{I}(z)} = 0\}$$

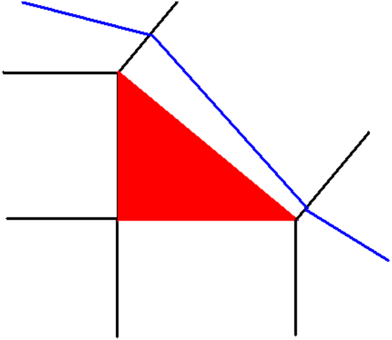


$C = \{z | Bz \geq b\}, F(z) = Mz + q$



The PATH algorithm

- Start in cell that has interior (face is an extreme point)
- 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 $\mathrm{rec}\mathcal{C}$ larger).
- Data generation
  - M is an $n \times n$ symmetric positive definite/indefinite matrix.
  - A has m randomly generated bounded inequality constraints.

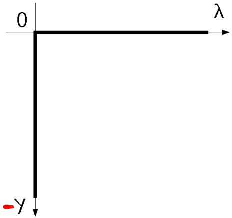
(m, n)	PathAVI		PATH		% negative eigenvalues
	status	# iterations	status	# iterations	
(180,60)	S	55	S	72	0
(180,60)	S	45	S	306	20
(180,60)	S	2	F	9616	60
(180,60)	S	1	F	10981	80
(360,120)	S	124	S	267	0
(360,120)	S	55	S	1095	20
(360,120)	S	2	F	10020	60
(360,120)	S	1	F	7988	80

Complementarity Systems (DVI)

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

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

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

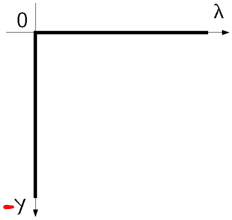


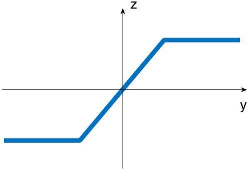
Complementarity Systems (DVI)

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

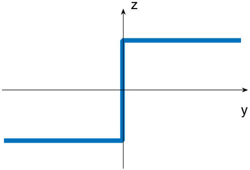
$$y(t) = h(x(t), z(t))$$

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

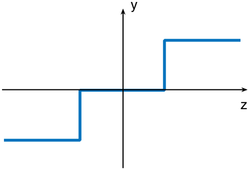




saturation



relay



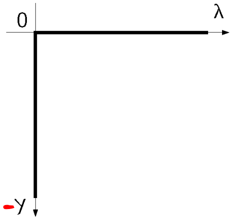
relay with dead zone

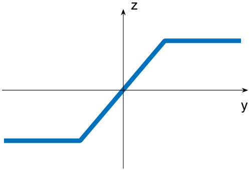
Complementarity Systems (DVI)

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

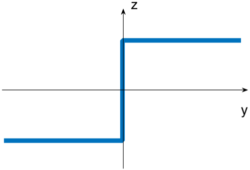
$$y(t) = h(x(t), z(t))$$

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

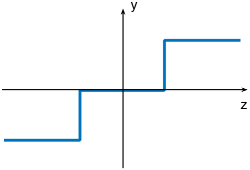




saturation



relay



relay with dead zone

Bimatrix Games: Golden Balls

- VI can be used to formulate many standard problem instances corresponding to special choices of M and $\mathcal{C}$.
- Nash game: two players have I and J pure strategies.
- p and q (strategy probabilities) belong to unit simplex $\triangle_I$ and $\triangle_J$ respectively.
- Payoff matrices $A \in R^{J \times I}$ and $B \in R^{I \times J}$, where $A_{j,i}$ is the profit received by the first player if strategy i is selected by the first player and j by the second, etc.
- The expected profit for the first and the second players are $q^T A p$ and $p^T B q$ respectively.
- A Nash equilibrium is reached by the pair of strategies (p^*, q^*) if and only if

$$p^* \in \arg \min_{p \in \triangle_I} \langle A q^*, p \rangle \quad \text{and} \quad q^* \in \arg \min_{q \in \triangle_J} \langle B^T p^*, q \rangle$$

Formulation using complementarity

The optimality conditions for the above problems are:

$$-Aq^* \in N_{\triangle_I}(p^*) \text{ and } -B^T p^* \in N_{\triangle_J}(q^*)$$

Therefore the corresponding VI is affine and can be written as:

$$0 \in \begin{bmatrix} 0 & A \\ B^T & 0 \end{bmatrix} \begin{bmatrix} p \\ q \end{bmatrix} + N_{\triangle_I \times \triangle_J}(\begin{bmatrix} p \\ q \end{bmatrix}). \tag{2}$$

What is EMP?

Annotates existing equations/variables/models for modeler to provide/define additional structure

- equilibrium
- vi (agents can solve min/max/vi)
- bilevel (reformulate as MPEC, or as SOCP)
- dualvar (use multipliers from one agent as variables for another)
- QS functions (both in objectives and constraints)

- Currently available within GAMS (full license available to course participants until March X, 2017 - contact me!)
- Some solution algorithms implemented in modeling system - limitations on size, decomposition and advanced algorithms
- QS extensions to Moreau-Yoshida regularization, compositions, composite optimization