International Conference on Continuous Optimization
Summer School, 1 August 2004
Rensselaer Polytechnic Institute

Tutorial on
Equilibrium Programming

Danny Ralph
Judge Institute of Management, Cambridge University

Contents

Part One: Equilibrium models and techniques based on nonsmooth equations

Part Two: Overview of Mathematical Programs with Equilibrium Constraints

This tutorial will NOT cover:

- Interior-point approaches to equilibrium problems
- Equilibrium (or optimisation) over positive semi-definite matrices or conic problems more generally

Part One: Equilibrium models and techniques based on nonsmooth equations

1. **Equilibrium formulations and applications**
2. **Reformulation as nonsmooth equations**
3. **Computation with normal equations for NCPs**
4. **Active research topics**
5. **Part One summary**

Unreferenced material largely in two volume opus on last 25 years of analysis & algorithms

Facchinei and Pang, Springer 2003
Finite-Dimensional Variational Inequalities and Complementarity Problems

1 Equilibrium formulations and applications

1.1 Basic equilibrium formulations

1.2 Applications

1.1 Basic equilibrium formulations

Nonlinear optimisation yields equilibrium problems:

Optimisation	**Equilibrium** = 1st-order conditions
$\min_{y} \phi(y)$	$0 = \nabla\phi(y)$, **"square" system of equations**
$\min_{y\geq 0} \phi(y)$	$0 \leq y \perp \nabla\phi(y) \geq 0$ where $\perp$ means orthogonal **Nonlinear Complementarity Prob., NCP**
$\min_{y} \phi(y)$ subj. to $g(y) \leq 0$ $h(y) = 0$	$\nabla\phi(y) + \nabla g(y)^\top\lambda + \nabla h(y)^\top\mu = 0$ $0 \geq g(y) \perp \lambda \geq 0$ $0 = h(y)$ **Mixed nonlinear Complementarity Prob., MCP**

Optimisation over **geometric** (abstract) constraints:

Optimisation	**Equilibrium** = 1st-order conditions
$\min_{\mathbf{y}\in\mathbf{C}} \phi(y)$	$y \in C$
	for all $c \in C$, $0 \leq \langle \nabla\phi(y), c - y\rangle$
	Variational inequality, VI

General VI:

	Equilibrium
	$y \in C$ for all $c \in C$, $0 \leq \langle \mathbf{F(y)}, c - y \rangle$ **Variational inequality, VI**

In general equilibrium problems: CPs and VIs are specified using a nonlinear vector function F(y) instead of $\nabla\phi(y)$, and constraints that are functional or geometric

If $m \times m$ Jacobian matrix $\nabla F(y)$ is **symmetric** for all y, then equilibrium system is stationary condition for optimisation problem.

But an equilibrium model need **not** be the stationary conditions of an optimization problem, e.g. Nash games.

1.2 Applications

We'll stick to finite dimensional problems, or discretisations of infinite dimensional problems. Note there is an accelerating interest in understanding infinite dimensional applications.

Main areas are economics and mechanics.

Economics

- General equilibrium: models of competitive markets
- Taxation and subsidies
- Spatial price models of distributed markets
- Nash games

Engineering

- Contact problems in mechanics, e.g. how does an elastic membrane wrap around a rigid object? C.f. classical obstacle problem, infinite dimensional.

- Contact problems with friction, e.g. how does a stack of blocks collapse if toppled?

- Structural mechanics, e.g. what is the shape of a bridge when a fleet of loaded trucks crosses?

- Traffic equilibria in networks, e.g. what will be the effect of introducing traffic lights at an intersection?

We'll give further motivation in an application in economics and another in mechanics below.

Cournot-Nash game

Consider a market for a single good with $i = 1, \ldots, N$ competitive producers; each decides what quantity q_i of this good to manufacture. "Nash" game is competitive, not collusive; players' strategies are their quantities, hence game termed "Cournot".

In this market, the price p of good depends on total output $Q = \sum q_i$: $p = p(Q)$.

Writing $q = (q_1, \ldots, q_N)$, player i will choose q_i to maximize profit

$$\pi_i(q_i) := p(Q)q_i - c_i(q_i)$$

where the latter term is the cost of manufacture.

Cournot-Nash game

Player i's stationary condition is

$$0 = \pi_i'(q_i) \;=\; p'(Q)q_i + p(Q) - c_i'(q_i).$$

The equilibrium system, of all players' stationary conditions, is $F(q) = 0 \in \mathbb{R}^N$ where $F_i(q) = \pi_i'(q_i)$ for each i.

Get asymmetry if pricing is nonlinear ($p''(Q) \neq 0$) and players not identical (i.e. expect $q_{\mathbf{i}} \neq q_{\mathbf{j}}$):

$$\begin{aligned} \partial F_{\mathbf{i}}/\partial q_{\mathbf{j}} \;&=\; p''(Q)q_{\mathbf{i}} + p'(Q) \\ &\neq\; p''(Q)q_{\mathbf{j}} + p'(Q) \;=\; \partial F_{\mathbf{j}}/\partial q_{\mathbf{i}}. \end{aligned}$$

So cannot expect to find equilibrium via optimisation.

Analysis of elasto-plastic structures

coming out of the Italian school in this research in 1960s and 70s.

Deformation (e.g. of a dam wall) is elastic - and reversible - for small loads.

Plastic deformation occurs after “yield limit” $\hat{R}$ is exceeded, modelled by plastic multiplier λ.

Equilibrium	$\hat{F} = \hat{C}^\top Q$	force vs. stress
Compatibility	$q = \hat{C} u$	strain vs. displacement

$$\text{Constitutive law} \quad \begin{cases} q = \epsilon + p & \text{strain} = \text{elastic} + \text{plastic strain} \\ Q = \hat{S}\epsilon & \text{stress v.s. elastic strain} \\ p = \hat{N}\lambda & \text{plastic strain vs. plastic multipliers} \\ \phi = \hat{\mathbf{N}}^\top \mathbf{Q} - \hat{\mathbf{H}}\lambda - \hat{\mathbf{R}} \leq \mathbf{0}, \quad \lambda \geq \mathbf{0}, \quad \phi^\top \lambda = \mathbf{0} \end{cases}$$

This is a mixed **l**inear **c**omplementarity **p**roblem, LCP

2 Reformulation as nonsmooth equations

2.1 Equilibrium problems as equations

2.2 Why bother to reformulate anyway?

2.1 Equilibrium problems as equations

Let $y \in \mathbb{R}^m$ and $F : \mathbb{R}^m \to \mathbb{R}^m$ be smooth, say C^2 or $C^{1,1}$.

Focus on NCP $\mathbf{0} \le \mathbf{y} \;\perp\; \mathbf{F}(\mathbf{y}) \ge \mathbf{0}$ for notational simplicity.
Note NCP equivalent to componentwise complementarity,

$$y_i, F_i(y) \ge 0, \quad y_i F_i(y) = 0 \quad \text{for each } i = 1, \ldots, m.$$

Could also consider $\mathrm{VI}(F, C)$ where C is any closed convex set.

- Easiest case: C is polyhedral convex (e.g. for NCP, $C = \mathbb{R}^n_+$)
- More generally, local analysis well developed for nonconvex C defined by smooth constraints

  $$g(x) \le 0 \in \mathbb{R}^k, \; h(x) = 0 \in \mathbb{R}^\ell$$

  satisfying a Constraint Qualification, CQ.

  CQs ensure, first and foremost, that solution of $\mathrm{VI}(F, C)$ has multipliers for constraints $\Longrightarrow$ can convert VI to mixed CP.

Aside on Constraint Qualifications at a feasible point x:

LICQ. Requires linear independence of gradients of active constraints, ensures uniqueness of multipliers.

MFCQ. Mangasarian-Fromowitz CQ requires LICQ on equality constraints ($\nabla h(x)$ full rank) and existence of an interior-pointing direction d: $\nabla h(x)d = 0$ and $\nabla g_i(x)^\top d < 0$ for active indices i. MFCQ equivalent to stability of C, near x, under perturbations of g and h, also equivalent to having nonempty, bounded set of multipliers.

CRCQ. Requires any list of constraints that are active at x to have Jacobian of same rank for all nearby feasible points. Seems to characterise polyhedral-like behaviour of nonlinear constraints [Kuntz-Scholtes 95]. Directional differentiability of solutions of parametric VIs may use Constant Rank CQ instead of LICQ.

Reformulate NCP $\mathbf{0} \le \mathbf{y} \perp \mathbf{F}(\mathbf{y}) \ge \mathbf{0}$ as **nonsmooth** equation.

$0 = \min\{y,\ F(y)\}$, componentwise,

min equation

$0 = F_+(z) := F(z_+) + z - z_+$

where $z_+ := \max\{0, z\}$, componentwise ($y := z_+$ will solve NCP)

normal equation

$\Phi(y, F(y)) = 0$, system of m equations where

$\Phi_i(y, F(y)) := \phi(y_i, F_i(y))$ is "NCP function"

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

These nonsmooth systems all **square**, having m var. and equations.
Min and normal equations are piecewise smooth; FB is semismooth.

Look at piecewise structure of F_+

Consider any orthant in $\mathbb{R}^m$, given by all z with

$$z_I \geq 0_I \quad \text{and} \quad z_J \leq 0_J$$

where $I \cup J$ partition $\{1, \ldots, m\}$. For such z,

$$F_+(z) \; = \; F(z_I, 0_J) + (0_I, z_J) \quad \ldots \text{ a smooth mapping.}$$

So F_+ is a different smooth mapping on each orthant, and also continuous, hence piecewise smooth, PC^1

C.f. solution y^* of NCP, whose associated active sets
$I = \{i \,:\, y_i^* > 0\}$ and $J =$ complement of I give

$$\begin{aligned} y_I^* \;&\geq\; 0 \;=\; F_I(y_I^*, 0_J) \\ y_J^* \;&=\; 0 \;\leq\; F_J(y_I^*, 0_J) \end{aligned}$$

Normal equation solved by $z_I^* = y_I^*$ and $z_J^* = -F_J(y_I^*, 0_J)$.

2.2 Why bother to reformulate anyway?

First equation reformulations aren't new, are useful!

Consider VI(F, C) where $F : \mathbb{R}^m \to \mathbb{R}^m$ is continuous, C nonempty closed convex set in $\mathbb{R}^m$. This is equivalent to finding a fixed point

$$y = \pi_C(y - F(y))$$

where $\pi_C(z) :=$ nearest point in C to z.
When C is also bounded, Brouwer's fixed point theorem ensures existence of a solution.

Note **NCP** equivalent to VI$(F, \mathbb{R}^n_+)$, and $\pi_{\mathbb{R}^n_+}(z) = z_+$.
Here, **fixed point formulation** with $C = \mathbb{R}^m_+$ is

$$y = (y - F(y))_+ \quad (= y - \min\{y, F(y)\}),$$

$\longrightarrow$ restatement of **min equation**.

Second, smooth calculus **motivates** tools for equilibrium systems.

Take **smooth system** $F(y) = 0$ of m vars. & equations. Suppose y^* is solution and $m \times m$ Jacobian $\nabla F(y^*)$ is invertible.

I. Newton's method: Given any y near y^* can find y^* quickly,

1. Let y' solve linearised system: $0 = F(y) + \nabla F(y)(y' - y)$.
2. Update $y := y'$ and repeat from step 1.

If F is modelled by choice of parameter $p = p^*$ as $F(y) = \Phi(y, p^*)$:
II. Implicit function theorem: (a) For p near p^* there is unique $y = y(p)$ near y^*, with $y(p^*) = y^*$, that solves $\Phi(y, p) = 0$;
(b) $y(p)$ is smooth and $\nabla y(p^*)$ can be found by solving

$$0 = \nabla F(y^*)\nabla y(p^*) + \nabla_p \Phi(y^*, p^*).$$

III. Parameter estimation: Suppose $\hat{y}$ measured empirically and $\Phi(\hat{y}, p^*) \neq 0$. Then update p via nonlinear least squares:

$$\min_{y,p} \|y - \hat{y}\|_2{}^2 \quad \text{subject to} \quad \Phi(y, p) = 0$$

Just to mention semismooth Newton methods ...

Although we focus, in §3, on piecewise linear-based solution techniques, Newton methods for semismooth systems, particularly FB reformulation of NCP, need to be mentioned.

Generalised Jacobian for FB mapping $\Phi(y)$. Although Φ is nonsmooth, you can easily give a (Clarke-like) Jacobian substitute

$$\partial\Phi(y) \ := \ \left\{ M \in \mathbb{R}^{m\times m} \text{ with } i\text{th row in } \ \partial_B\phi(y_i, F_i)^\top \nabla(y_i, F_i) \right\}$$

where F_i means $F_i(y)$ and

$$\partial_B\phi(a,b) = \begin{cases} (1,1) - (a,b)/\|(a,b)\|_2 & \text{if } (a,b) \neq (0,0), \\ (1,1) - \text{ Euclidean unit circle} & \text{otherwise.} \end{cases}$$

Newton's method for FB system $\Phi(y) = 0$. Given y, and (invertible) $M^y \in \partial\Phi(y)$, next iterate y' solves "linearised" system

$$0 \ = \ \Phi(y) + M^y(y' - y).$$

FB mapping has two remarkable properties, where F is $\mathrm{C}^{1,1}$.

1. (Strong) Semismoothness. Suppose y^* is given. There is $C = C^{y^*} > 0$ s.t. for any y near y^*, any $M^y \in \partial\Phi(y)$,

$$\|[\Phi(y) + M^y(y^* - y)] - \Phi(y^*)\| \leq C\|y - y^*\|^2$$

Linearised system is good approximation, within $O(\|y - y^*\|^2)$!

Leads to **quadratic convergence of semismooth Newton method** to y^* if $\Phi(y^*) = 0$, $\partial\Phi(y^*)$ contains only invertible matrices, and first iterate is near y^*.

2. Residual squared is smooth. $\theta(y) := \|\Phi(y)\|_2{}^2$ is $\mathrm{C}^{1,1}$.

This leads to **globally convergent methods**, e.g. given y, apply linesearch in either semismooth Newton direction $y' - y$, or steepest descent direction $-\nabla\theta(y)$ if M^y is singular or nearly so.

3 Computation with normal equations for NCPs

3.1 Preamble: Invertibility of M_+

3.2 Newton's method for $F_+(z) = 0$

3.3 An implicit function theorem

3.4 Parameter estimation for NCPs

3.1 Preamble: Invertibility of M_+

Suppose $F(y) = My + q$ for given $M \in \mathbb{R}^{m\times m}$, $b \in \mathbb{R}^m$.

Q: When is $\min\{y, My + q\}$ invertible?
What about $F_+(z) := Mz_+ + q + z - z_+$?

A: Well known result that M must be a P-matrix: all principle minors have positive determinants. (E.g. positive definite M gives invertibility.)
Easy to show that P-matrix $\Rightarrow$ existence of sol.
Hard to show uniqueness.

More subtle Q: What characterises invertibility **near** given y^* ?

A: Well known result based on fixing $(My + q)_i = 0$ if $i \in I := \{j : y_j^* > 0\}$, and $y_i = 0$ if $i \in J := \{j : (My^* + q)_j > 0\}$.
Let K be complement of $I \cup J$; require submatrix M_{II} invertible s.t. $M_{KK} - M_{KI}M_{II}^{-1}M_{IK}$ is P-matrix.

Solving piecewise linear systems

Lemke's method for solving LCPs is easily adapted to the normal equation $0 = M_+(z) + q$, and indeed to general piecewise linear systems [Eaves 76]

Use a kind of active-set approach, trying to identify $\{i : z_i^* > 0\}$ at solution z^*:

- perform a sequence of pivots, from one linear piece of the system to the next
- carry out a rank-1 update of a linear system solve each time
- (for invertible mapping) terminate after finitely many steps with solution

Interior-point methods provide polynomial-time solution for LCPs with positive definite (implies P-)matrices and more general classes of matrices.

3.2 Newton's method for $F_+(z) = 0$

We'll develop normal equation viewpoint in rest of §3. The min, FB or other semismooth reformulations provide alternatives. Restrict discussion to normal mapping F_+ for NCP, for simplicity.

Newton's Method.

1. "Linearise" system $0 = F_+(z) := F(z_+) + z - z_+$ about current point z and solve to find z'.
I.e. linearize F about z_+ but don't mess with $(\cdot)_+$: Let $M^z := \nabla F(z_+)$, $q^z := F(z_+) - \nabla F(z_+)z_+$, then solve for z',

$$0 = M^z_+(z') + q^z \ .$$

(can equivalently solve LCP or $\min\{y', M^z y' + q^z\} = 0$, then convert y' to z')

2. Update $z := z'$ and repeat 1.

When does Newton's method work?

Look at **smooth ($\mathbf{C}^{1,1}$) system** $F(y) = 0$ [Ortega-Rheinboldt 70].

Banach perturbation lemma.

If $M \in \mathbb{R}^{m \times m}$ is invertible and $\epsilon \in (0, 1)$ then

- so is any $\hat{M}$ within distance $\epsilon / \|M^{-1}\|$ of M,
- $\|\hat{M}^{-1}\| \leq \|M^{-1}\| / (1 - \epsilon)$

Apply to case when $F(y^*) = 0$ and $M := \nabla F(y^*)$ invertible. Get for y near y^* that $\nabla F(y)$ invertible with $\|\nabla F(y)^{-1}\|$ bounded.

For y near y^*, Newton step $y' = y - \nabla F(y)^{-1} F(y)$ exists, and satisfies

$$\begin{aligned}
\mathbf{y' - y^*} &= (y' - y) + (y - y^*) \\
&= -\nabla F(y)^{-1}\left[F(y) + \nabla F(y)(y^* - y)\right] \\
&= -\nabla F(y)^{-1}\left[F(y^*) + O(\|y - y^*\|^2)\right] \\
&= -\nabla F(y)^{-1}\left[O(\|y - y^*\|^2)\right] = \mathbf{O(\|y - y^*\|^2)}
\end{aligned}$$

Quadratic convergence!

because

$\|\nabla F(y)^{-1}\|$ is bounded, and

linearization gives good approximation:

$$F(y) + \nabla F(y)(y^* - y) \;=\; F(y^*) + O(\|y - y^*\|^2).$$

When does Newton's method work?

Analysis for **nonsmooth normal equations** with $\mathrm{C}^{1,1}$ mapping F,

$$0 = \mathbf{F}_+(\mathbf{z}) := \mathbf{F}(\mathbf{z}_+) + \mathbf{z} - \mathbf{z}_+.$$

Simplifying (from local to global invertibility) here:

Banach-perturbation-type lemma.
If M_+ invertible, where $M \in \mathbb{R}^{m\times m}$, then inverse M_+^{-1} has Lipschitz constant denoted $\|M_+^{-1}\|$.
And, for $\epsilon \in (0,1)$,

- any $\hat{M}$ within distance $\epsilon/\|M_+^{-1}\|$ of M gives invertible $\hat{M}_+$;
- $\hat{M}_+^{-1}$ has Lipschitz constant $\|\hat{M}_+^{-1}\| \le \|M_+^{-1}\|/(1-\epsilon)$

Apply to case when $F_+(z^*) = 0$, $M := \nabla F(z_+^*)$ and M_+ is invertible.

Given z near z^*, $M^z := \nabla F(z_+)$ and $q^z := [F(z_+) - \nabla F(z_+) z_+]$,
Newton step is solution $z' := {M_+^z}^{-1}(-q^z)$.

Analog of analysis for smooth case gives

$$
\begin{aligned}
z' - z^* &= {M_+^z}^{-1}[-q^z] - {M_+^z}^{-1}[M_+^z(z^*)], \\
\|\mathbf{z}' - \mathbf{z}^*\| &\le \|{M_+^z}^{-1}\| \| - q^z - M_+^z(z^*)\| \\
&= \|{M_+^z}^{-1}\| \big\| F_+(z^*) + O(\|z - z^*\|^2) \big\| \\
&= \mathbf{O}(\|\mathbf{z} - \mathbf{z}^*\|^{\mathbf{2}})
\end{aligned}
$$

Quadratic convergence!

because, similar to smooth case,

$\|{M_+^z}^{-1}\|$ is bounded, and

linearization gives good approximation:

$$M_+^z(z^*) + q^z = F_+(z^*) + O(\|z - z^*\|^2).$$

Extension i. A stronger convergence result, of **Newton-Kantorovich** type: **If**

- linearization P_z at initial iterate z is invertible,
- $\|P_z^{-1}\|$ is "not too large" relative to initial residual $\|F_+(z)\|$ and Lipschitz constant of ∇F near z

then

- there exists a solution z^*, near z, of normal equation
- starting from z, Newton's method generates sequence that quadratically converges to z^*.

Extension ii. Newton's method can be globalised for robustness.

Idea of **line search for smooth equations** $F(y) = 0$:
Given y, take Newton iterate y', observe for $t \in (0, 1]$ that

$$F(y + t(y' - y)) = (1 - t)F(y) + O(t^2).$$

Choose "large" $t \leq 1$ s.t. $\|F(y + t(y' - y))\| < \|F(y)\|$ and
next iterate $:= y + t(y' - y)$

For **normal equations**, use "path search", c.f. PATH code.
Given z, calculate **Newton path** $z(t)$ by solving parametric linearised system:

$$M_+^z(z(t)) + q^z \; = \; (1 - t)F_+(z) \quad \text{for } t \in [0, 1].$$

Analogous to smooth case, get

$$F_+(z(t)) \; = \; (1 - t)F_+(z) \; + \; O(t^2).$$

Choose "large" $t \leq 1$ s.t. $\|F_+(z(t))\| < \|F_+(z)\|$ and $z' := z(t)$.

3.3 An implicit function theorem

It follows from above that

- if $F(y) = \Phi(y, p^*)$ for parameter p^*
- if normal map $F_+(z) := F(z_+) + z - z_+$ is invertible near a zero z^*

then for p near p^* there is a solution $z = z(p)$, unique near z^* and with $z(p^*) = z^*$, of parametric normal equation

$$0 \;=\; \Phi(z_+, p) + z - z_+$$

And $y(p) := z(p)_+$ is locally unique solution of parametric NCP.

Sensitivity: $z(p)$ is directionally differentiable, and, for small perturbations dp, can calculate $z'(p^*; dp) = z - z^*$ by solving for z in parametric linearised system (or, equivalently, LCP)

$$0 \;=\; M_+(z) + q + \nabla_p \Phi(z_+^*, p^*)dp$$

where $M := \nabla F(z_+^*)$ and $q := F(z_+^*) - M z_+^*$

3.4 Parameter estimation for NCPs

Suppose

- $F(y) = \Phi(y, p^*)$ for some parameter vector p^*
- we have measured, empically, NCP solution as $\hat{y}$,
- but $y = \hat{y}$ does not quite solve the NCP: $0 \le y \perp F(y) \ge 0$.

Apply least squares idea to estimate p

$$\begin{array}{cl} \min\limits_{y,p} & \|y - \hat{y}\|_2{}^2 \\ \text{subject to} & y \text{ solves NCP: } 0 \le y \perp \Phi(y, p) \ge 0. \end{array}$$

... a Mathematical Program with Equilibrium Constraints, MPEC

4 Active research topics

Do a GOOGLE search:

VIs over positive semidefinite matrix variables

- Interior-point methods for SDP!
- Semismooth Newton methods for SDP

Dynamic variational inequalities

Stochastic variational inequalities

Infinite dimensional variational inequalities, PDE

5 PART ONE Summary

- Equilibrium problems in the form of complementarity problems and variational inequalities are all around us

- Many lessons from smooth systems of equations can be carried over — directly — to nonsmooth reformulations

- Parameter estimation for equilibrium problems gives MPECs.

  - The gap between nonlinear programs (NLPs) and MPECs is still being explored, **stay tuned for review of MPEC ...**

General references on equilibrium modelling etc.

Lanchester prize winner

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

Excellent review of econ and eng applications

M.C. Ferris and J.S. Pang, SIAM Review 1997

A classic: infinite dimensional VIs in applied maths (PDE)

D. Kinderlehrer and G. Stampacchia, *An Introduction to Variational Inequalities and Their Applications*, SIAM 2000.

Well known monograph in social sciences

A. Nagurney, *Network Economics: A Variational Inequality Approach* (Revised second edition), Kluwer 1999.

Another well regarded monograph, an optimisation view

M. Patriksson,*Nonlinear Programming and Variational Inequality Problems: A Unified Approach*, Kluwer 1998.

PART TWO: Overview of MPECs

6. Introduction to MPEC

7. MPEC and NLP

8. Active research topics

9. Conclusion

6 Introduction to MPEC

6.1 Milestones

6.2 Formulation

6.3 Where do MPECs come from?

6.4 Are MPECs typical optimization problems?

6.1 Milestones†

MPEC models go back to 1930s (von Stackelberg)

Pre 96: Focus on heuristic algorithms for global bilevel optimisation

1996 onwards:

- Explicit efforts to create academic subject, e.g. monographs [Luo-Pang-R 96], [Outrata-Kocvara-Zowe 98], c.f. bilevel programming: [Bard 98], [Dempe 02]
- Local optimization of MPECs in applications
- Plethora of algorithms using decomposition or nonsmooth/smoothing reformulations

2001ish: Standard NLP methods make an impact, MacMPEC test set (Leyffer)

† Adapted from Stefan Scholtes

6.2 Formulation

Mathematical Programs with Equilibrium Constraints have the general form

$$\min_{\mathbf{x},\mathbf{y}} \quad f(\mathbf{x},\mathbf{y}) \text{ subject to}$$

$$(\mathbf{x},\mathbf{y}) \in Z \quad = \quad \text{standard region, e.g. polyhedron}$$

$$\mathbf{y} \text{ solves} \quad \left\{ \begin{array}{l} \text{optimization} \\ \text{equilibrium} \end{array} \right\} \textbf{problem} \text{ that depends on } \mathbf{x}$$

where
- $\mathbf{x} \in \mathbb{R}^n$ is **leader** or **design** or **control** vector,
- $\mathbf{y} \in \mathbb{R}^m$ is **follower** or **response** or **state** vector,
- $f : \mathbb{R}^{n+m} \to \mathbb{R}$ is smooth, and
- lower-level optimization or equilibrium problem is smooth.

Focus, later, on Complementarity Constraints (CC) for simplicity.

$$\begin{array}{lll} \min\limits_{x,y} & f(x,y) \text{ subject to} & \\ & (x,y) \in Z & \text{standard region, e.g. polyhedron} \\ & 0 \leq F(x,y), y \quad 0 = y^\top F(x,y) & \end{array}$$

where $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$, and $F : \mathbb{R}^{n+m} \to \mathbb{R}^m$ is smooth.

May refer to MPCC rather than MPEC.

More generally, equilibrium problems come from Variational Inequalities, famous in economic equilibria models, traffic equilibria, even partial differential equations.

6.3 Where do MPECs come from?

Equilibrium model. E.g. *Traffic network*: Given network, origin-destination volumes, function describing length of journey, find link flows at which every user is on a shortest path.

Design. Set the parameters of an equilibrium model to achieve an objective. E.g., *Traffic network:*
Set tolls on designated links to maximise revenue while bounding worst case congestion of equilibrium traffic flow

Parameter Identification. Corresponding to an equilibrium model for a real system is an inverse problem: find parameters of the model for which the model equilibrium matches observed (measured, actual) equilibrium. E.g., *Traffic network:*
Given empirical link flows (assumed to be equilibrium flows), find origin-destination flows.

6.3 Where do MPECs come from?

Equilibrium model. E.g. *Structural Analysis*: given topology of truss = network of beams, material properties of each beam, and the load it supports, find its shape, stresses, strains etc.

Design. Set the parameters of an equilibrium model to achieve an objective. E.g. *Structural Analysis*:
determine least volume (cost) of beams in a bridge while suppporting a given load

Parameter Identification. Corresponding to an equilibrium model for a real system is an inverse problem: find parameters of the model for which the model equilibrium matches observed (measured, actual) equilibrium. E.g. *Structural Analysis*:
given the load on, and displacement (shape) of a structure, find the properties of the materials used in construction ... which may change over time

Recall **LCP for elasto-plastic structures**

Plastic deformation occurs after "yield limit" $\hat{R}$ is exceeded, modelled by plastic multiplier λ.

Equilibrium	$\hat{F} = \hat{C}^\top Q$	force vs. stress
Compatibility	$q = \hat{C} u$	strain vs. displacement

$$\text{Constitutive law} \quad \begin{cases} q = \epsilon + p & \text{strain = elastic + plastic strain} \\ Q = \hat{S}\epsilon & \text{stress v.s. elastic strain} \\ p = \hat{N}\lambda & \text{plastic strain vs. plastic multipliers} \\ \phi = \hat{\mathbf{N}}^\top \mathbf{Q} - \hat{\mathbf{H}}\lambda - \hat{\mathbf{R}} \leq \mathbf{0}, \quad \lambda \geq \mathbf{0}, \quad \phi^\top \lambda = \mathbf{0} & \end{cases}$$

This is a mixed **l**inear **c**omplementarity **p**roblem, LCP

Identification problem of finding yield limits

Given • **LCP** model including $\hat{F}$ = loading, $\hat{H}$ = hardening,
but **not** $\hat{R}$ = yield limits,
• $\hat{u}$ = measured displacements of actual structure,

estimate yield limits R by solving the MPEC

$$\begin{array}{ll} \displaystyle\min_{\mathbf{R},\,(\mathbf{Q},\mathbf{u},\lambda)} & \|\mathbf{u}-\hat{\mathbf{u}}\|^2 \\ \text{subject to} & (\mathbf{Q},\mathbf{u},\lambda) \text{ solves the } \textbf{Mixed LCP}(\mathbf{R}), \\ & \text{bounds on } \mathbf{R}. \end{array}$$

Here $\mathbf{R}$ is the leader, $(\mathbf{Q},\mathbf{u},\lambda)$ is the follower = piecewise linear function of $\mathbf{R}$.

This is an **inverse parameter identification** problem.

6.4 Are MPECs typical optimization problems?

Looking ahead, **No.**

MPECs are unusually challenging nonlinear programs:

- have **nonconvex** polyhedral feasible sets even when all constraint functions are linear

- violate NLP constraint qualifications if complementarity between $y \geq 0$ and $F(x, y) \geq 0$ is written as dot product $y^\top F(x, y) = 0$

- nevertheless, seem amenable to NLP analysis and algorithms in many cases.

7 MPEC & NLP

7.1 Decomposition

7.2 B-stationarity and a linear independence constraint qualification

7.3 Sequential NLP methods

7.4 NLP methods

7.1 Decomposition

Focus on MPCC written as NLP:

$$\begin{array}{ll} \min\limits_{z \in \mathbb{R}^n} & f(z) \\ \text{subject to} & g(z) \leq 0 \\ & \mathbf{G(z) \geq 0, H(z) \geq 0} \\ & \mathbf{G(z)^\top H(z) = 0} \end{array} \qquad \text{(MPCC)}$$

where $f : \mathbb{R}^n \to \mathbb{R}$

$g : \mathbb{R}^n \to \mathbb{R}^p$

$G : \mathbb{R}^n \to \mathbb{R}^m,\ H : \mathbb{R}^n \to \mathbb{R}^m\ (n \geq m)$

are smooth, at least C^2. Omit upper-level equality constraints ...

Lower-level problem is parametric NCP when $z = (x, y)$, x is the leader/design/control, y is the follower/response/state, and $H(x, y) = y$. Then y solves

$$0 \leq G(x, y) \perp y \geq 0.$$

Example 1. Two-variable MPCC, $z = (x, y) \in \mathbb{R}^2$,

$$\min_{x,y \in \mathbb{R}} f(x, y) \text{ subject to } x, y \geq 0; \;\; x^\top y = 0.$$

Why is Example 1 **not** like a standard NLP?

Feasible set decomposes into

$$\begin{aligned} & \{(x, \mathbf{y}) : x = 0 \leq \mathbf{y}\} \\ \cup \quad & \{(\mathbf{x}, y) : \mathbf{x} \geq 0 = y\} \end{aligned}$$

which is **polyhedral nonconvex**; it consists of two "pieces" or "branches".

Standard difficulties for nonlinear programming (NLP) ideas applied to MPEC:

$\times$ The feasible set of is **nonconvex**, even when polyhedral

$\longrightarrow$ if x, y vectors in $\mathbb{R}^m$ then there are 2^m pieces

$\longrightarrow$ even checking stationarity is combinatorial

$\times$ Typical nonlinear **constraint qualifications**, e.g. Mangasarian-Fromowitz CQ, **fail** at *all* feasible points [Chen-Florian 94, Ye-Zhu-Zhu 97].

$\longrightarrow$ existence of Karush-Kuhn-Tucker (KKT) multipliers?

$\longrightarrow$ numerical stability?

In spite of these difficulties, decomposition into pieces is very useful under a linear independence condition ...

7.2 B-stationarity and a linear independence constraint qualification

Definition. A point feasible point of (MPCC) is B-stationary or piecewise stationary if it is stationary, i.e. has KKT multipliers, for each piece for which it is feasible.

Stationary conditions for Example 1

Let $(0,0)$ be local min, then it is "biactive", hence stationary for each of two pieces of (MPCC)

$$\min_{x,y} f(x,y) \quad \text{subject to} \quad x = 0 \ \leq \mathbf{y},$$

$$\min_{x,y} f(x,y) \quad \text{subject to} \quad \mathbf{x} \geq 0 = y.$$

That is, (0,0) is B-stationary.

If $x, y \in \mathbb{R}^m$ then B-stationarity of $(0,0) \in \mathbb{R}^m \times \mathbb{R}^m$ entails stationarity on 2^m pieces.

Continuing **Example 1** with local min (0,0):

These pieces of the MPEC share **same Lagrangian**, called **MPEC-Lagrangian**,

$$L(x, y; \mu, \nu) = f(x, y) \; - \; \tau^\top x \; - \; \mu^\top y$$

and **same KKT conditions** **excepting sign of multipliers** τ, μ.

So Lagrangian conditions for each piece are the same

$$\begin{aligned} 0 \quad &= \quad \nabla_x L \; = \; \nabla_x f(0,0) - \tau^* \\ 0 \quad &= \quad \nabla_y L \; = \; \nabla_y f(0,0) - \mu^* \end{aligned}$$

and, in fact, define **same multipliers** τ^*, μ^* for each piece. These **MPEC-multipliers** must satisfy sign conditions for each piece,

$$\tau^*, \; \mu^* \quad \geq \quad 0$$

Otherwise (0,0) cannot be B-stationary (or a local min).

Generally: if Lagrangian conditions define unique multipliers then can check B-stationarity by looking at sign of multipliers for biactive constraints [Luo-Pang-R 98, Scheel-Scholtes 00].

Stationary conditions are non-combinatorial !

Definition.
MPEC-LICQ says that **active constraints**, excluding complementarity equations ($x^\top y = 0$ in Example 1), have **linearly independent gradients**.

Obviously MPEC-LICQ ensures that Lagrangian conditions define unique multipliers, if there are any, hence stationarity is easy to check as above.

Check Example 1 at (0,0): active gradients — excluding the complementarity equation — are $(1,0)$ and (0,1), LI vectors.

MPEC-multipliers in decomposition active set approach

In Example 1, suppose $(0,0)$ is stationary for one piece. Look at KKT multipliers τ, μ corresponding to $x \geq 0$, $y \geq 0$ respectively.

- If $\tau < 0$ then, as in classical active set methods, there is descent direction with $x \geq 0$.
  $\Longrightarrow$ determine next iterate by one step of NLP method on piece of MPEC: $\min f(x,y)$ subject to $x \geq 0 = y$.
- If $\mu < 0$ then there is descent direction in piece $x = 0 \leq y$.
- Finally, if $\tau, \mu \geq 0$ then (0,0) is **B-stationary** point.

Globally convergent decomposition schemes [Stöhr-Scholtes 99, Stöhr 99, Fukushima-Tseng 02], recently [Giallombardo-R 04]. See also general approach in [Scholtes 03]. PSQP method closely related [Luo-Pang-R 96, 98], [Jiang-R 99, 02].

[Zhang-Liu 01] enumerates extreme rays, without MPEC-LICQ.

7.3 Sequential NLP methods

$$\text{Recall (MPCC)} \begin{cases} \min\limits_{z} & f(z) \\ \text{subject to} & g(z) \le 0 \\ & \mathbf{G(z), H(z) \ge 0}, \quad \mathbf{G(z)^\top H(z) = 0} \end{cases}$$

Sequential NLP:

- Embed (MPCC) into family (NLP_ε) indexed by scalar $\varepsilon > 0$ where, roughly speaking,
  - CQ holds for each (NLP_ε)
  - (NLP_ε) $\to$ (MPCC) as $\varepsilon \to 0_+$
- **Algorithm.**
  Given $\varepsilon^k > 0$, use NLP method to find a stationary point z^k of ($\mathrm{NLP}_{\varepsilon^k}$). Let $0 < \varepsilon^{k+1} < \varepsilon^k$, $k = k + 1$, and repeat.
- Expect good behaviour of limit points of $\{z^k\}$ if $\varepsilon^k \to 0_+$

Smoothing method goes back to [Facchinei-Jiang-Qi 99]. Use smoothed Fischer-Burmeister function ϕ^ε for small $\varepsilon > 0$:

$$\begin{array}{lll} \min\limits_z & f(z) & \\ \text{subject to} & g(z) \le 0 & \\ & \phi^\varepsilon(G_i(z), H_i(z)) = 0 \quad \text{for all } i \end{array} \qquad (S_\varepsilon)$$

where $\phi^\varepsilon(a, b) := a + b - \sqrt{a^2 + b^2 + \varepsilon}$ for any scalars a, b.

Regularization method, given small $\varepsilon > 0$:

$$\begin{array}{ll} \min\limits_z & f(z) \\ \text{subject to} & g(z) \le 0 \\ & G(z), H(z) \ge 0 \\ & G_i(z)H_i(z) \le \varepsilon \;\; (\text{or} = \varepsilon) \quad \text{for all } i \end{array} \qquad (R_\varepsilon)$$

Penalty method: Given small $\varepsilon > 0$, set penalty parameter $1/\varepsilon$ and move the complementarity equation into the objective:

$$\begin{array}{ll} \min\limits_{z} & f(z) + \frac{1}{\varepsilon} G(z)^T H(z) \\ \text{subject to} & g(z) \leq 0 \\ & G(z), H(z) \geq 0 \end{array} \qquad (P_\varepsilon)$$

Straightforward convergence analysis available for

Clarke-stationary points of smoothing method, using generalised gradients of active FB constraints

Globally optimal solutions, e.g.

If z^k is a **global min** of (P_{ε^k}) and $(z^k, \varepsilon^k) \to (\bar{z}, 0_+)$, then $\bar{z}$ is **global min** of (1).

Latter impractical since (P_{ε^k}) nonconvex $\Longrightarrow$ **local** solutions or **stationary** points of (P_{ε^k}) are computationally desirable.

Breakthrough for convergence of sequential NLP methods:

Smoothing. [Fukushima-Pang 00] smooth NLP family (S_ε)

Followed shortly by

Regularisation. [Scholtes 01] regularisation/relaxation (R_ε)

Penalty. [Hu-R 01] C^2 penalty family (P_ε)
[Huang-Yang-Zhu 01] F-B squared penalty term which is $C^{1,1}$ but not C^2.

Do Sequential NLP as above by "solving" $(\mathrm{NLP}_{\varepsilon^k})$

- z^{k+1} is KKT point of $(\mathrm{NLP}_{\varepsilon^k})$
- A **Weak Second Order Necessary Condition**, WSONC, is required at z^{k+1}
- Given MPEC-LICQ & nonzero multipliers of biactive complementarity constraints ("upper-level strict complementarity"), get **B-stationary** limit points of $\{z^k\}$

7.4 NLP methods

Sequential Quadratic Programming, SQP. Leyffer revisited NLP methods by demonstrating superlinear convergence of SQP codes `Filter` and `SNOPT` on MPECs, [Fletcher-Leyffer 02]

Local superlinear convergence analysis of SQP, e.g.

[Anitescu 03]
[Fletcher-Leyffer-R-Scholtes 02]
[Izmailov-Solodov 03]
[Wright 03]

Interior-point methods. Tailoring IP approaches to MPECs shown to be effective, e.g.

[Benson-Shanno-Vanderbei 02]
[DeMiguel-Friedlander-Nogales-Scholtes 03]
[Liu-Sun 01]
[Raghunathan-Biegler 03]

8 Active research topics

Do a GOOGLE search:

Stochastic MPEC

Global optimisation and bounding of MPEC

MPECs in infinite dimensions or Control of PDE

Bilevel games and EPECs

9 Conclusion

MPECs/MPCCs

- include
  - system design, optimising parameters in an equilibrium setting
  - parameter identification = inverse equilibrium problems, a large application area.
- are **not** standard **nonlinear programs**.
- but **are (locally) solvable** by some standard **NLP** techniques and codes
- require serious effort to understand computational performance