Recent Advances in Nonlinear Optimization

Jorge Nocedal

Optimization Technology Center
Northwestern University

Fields Institute, Feb 7, 2006,

New Applications, New Algorithms, New Software

The general nonlinear programming formulation covers many interesting applications

$$\begin{array}{ll} \min\limits_{x} & f(x) \\ \text{s.t.} & h_i(x) = 0,\ i \in E \\ & g_i(x) \geq 0,\ i \in I \end{array}$$

Why discuss algorithms?

Active research area; significant algorithmic advances

New applications

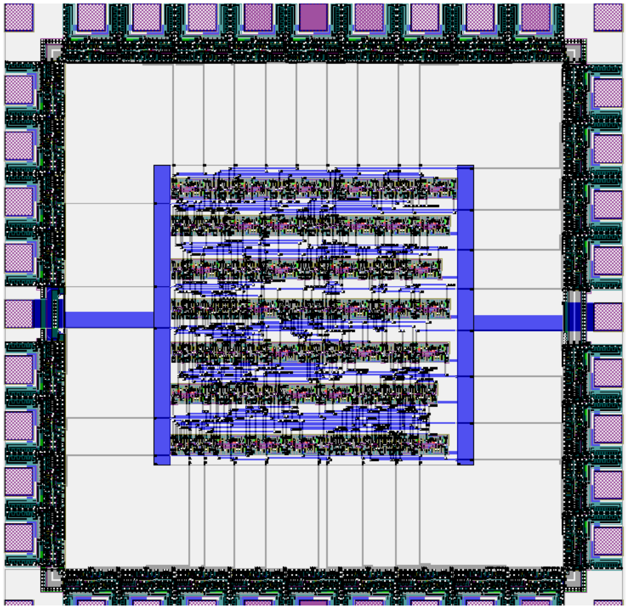
Solve larger problems, more difficult (degenerate), noisy

Interior-point, Active-set: dominate; Penalty approaches

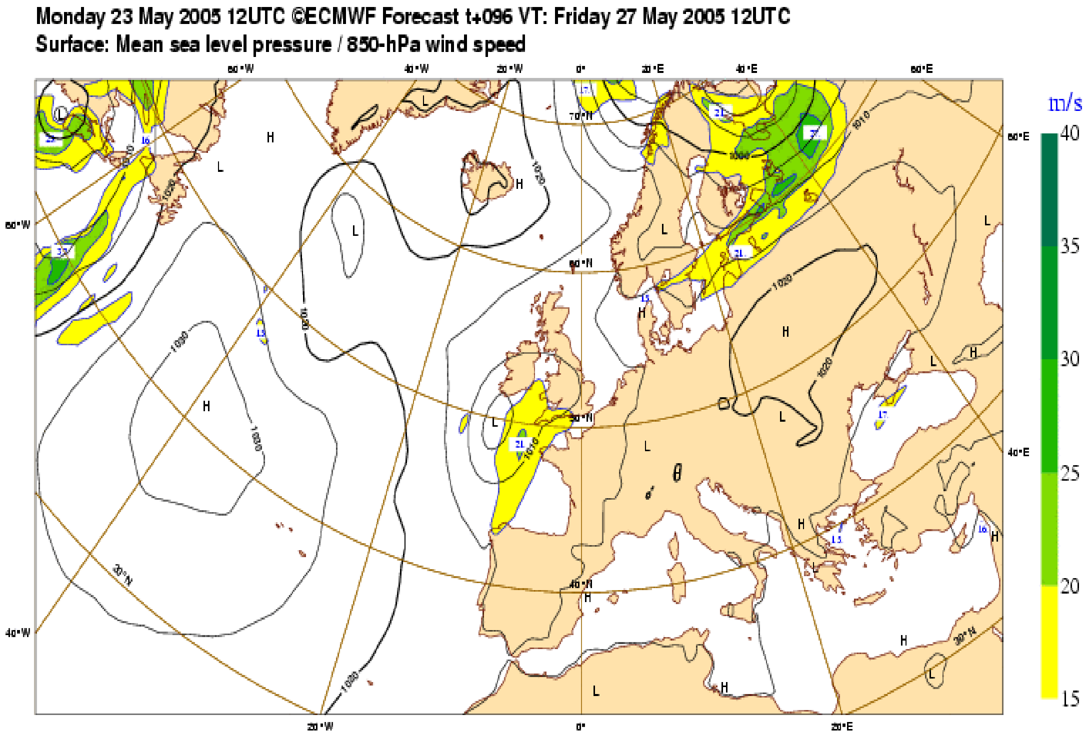
Outline

- Applications
- Software (KNITRO)
- New Optimization Methods
- Mathematical Foundations

Problem 1: Circuit simulation. Given a design of a computer chip (integrated circuit) find a way to resize elements to achieve optimal performance (power consumption)
Hundreds of thousands of elements. INTEL

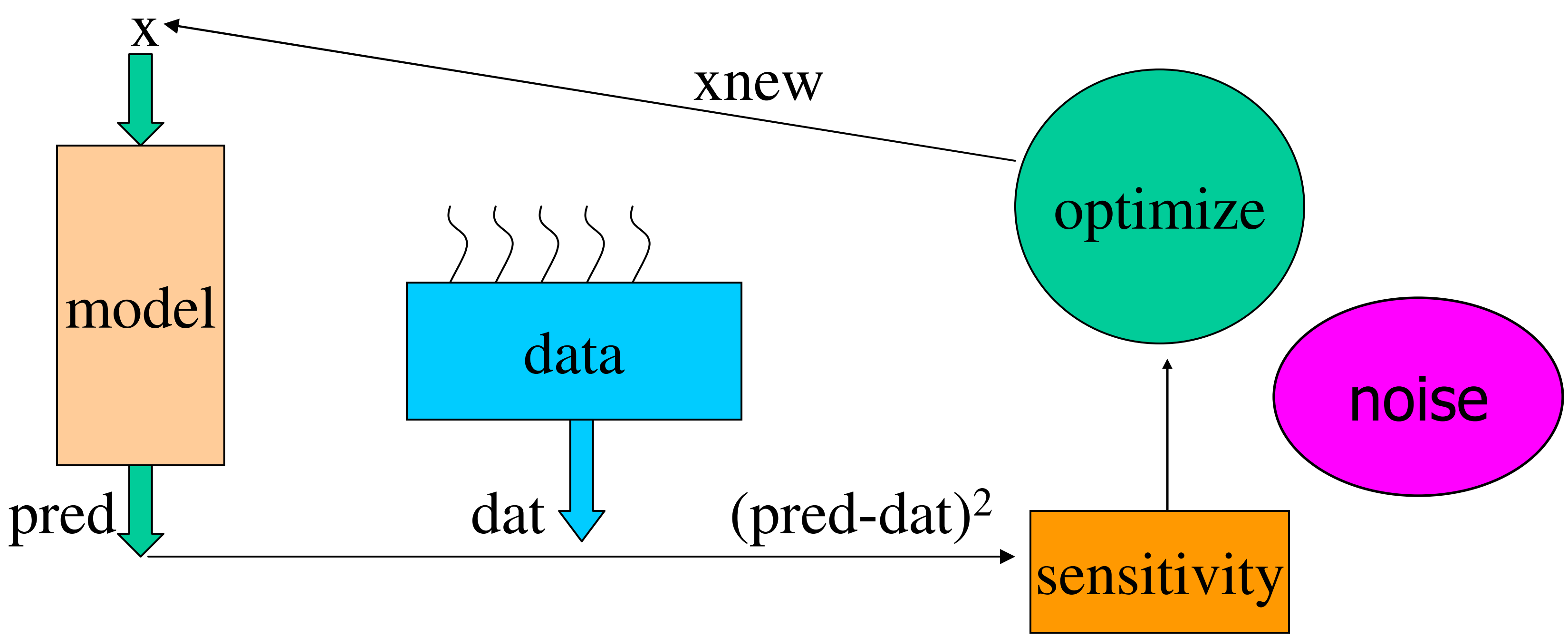


Problem 2: Guess initial conditions of the atmosphere at 1 million locations
so that predicted fluid flow matches observations during a 12-hour period. ECMWF



2 Cases: prototypical of many present/future applications

Mathematical model (circuit simulation)
Control it to produce the desired results



Cannot afford many attempts (20)
Lower resolution models

Some current research areas

- Larger and larger problems
- Noisy functions
- Integer/continuous variables
- Games (Nash, Stackelberg,…)
- Real time (warm starts)
- Degenerate problems (deficient geometry)

General NLP techniques preferred over specialized algorithms

Challenging geometry of feasible region

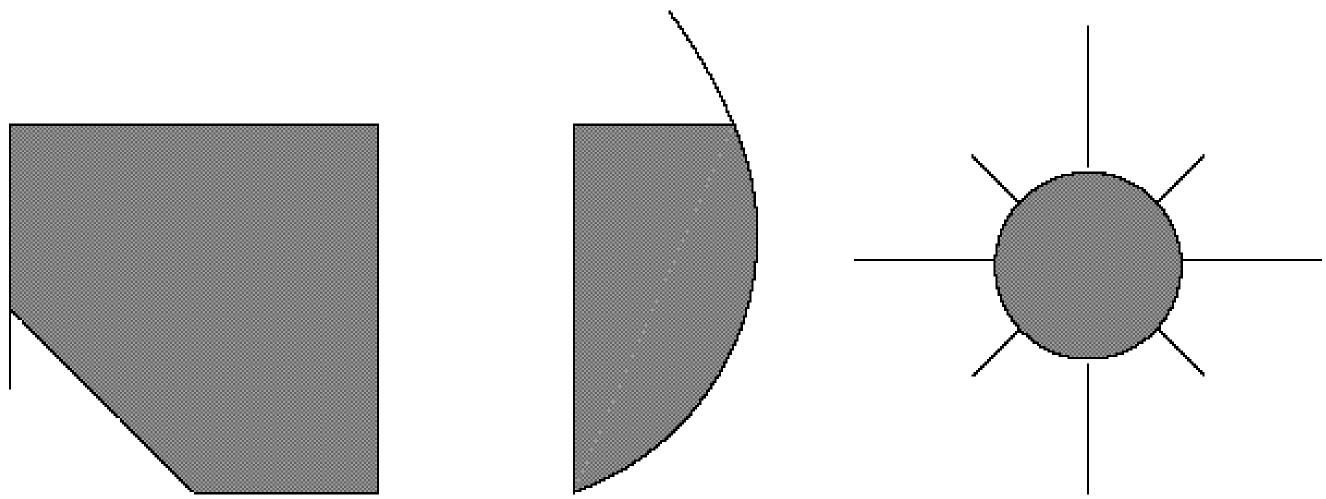


Figure 9.1: Feasible regions for switch-off problems.

Mathematical Difficulties

Many new areas of application are modeled as:

$$
\begin{array}{ll}
\min & f(x) \\
\text{s.t.} & g(x) \ge 0 \\
 & h(x) = 0 \\
 & \left.\begin{array}{r} x_1 x_2 = 0 \\ x_1, x_2 \ge 0 \end{array}\right\} \longrightarrow
\end{array}
$$

these constraints cause regularity to be lost

where does this occur? Optimality!

$$
g(x)\lambda = 0
$$

$$
g(x) \ge 0, \quad \lambda \ge 0
$$

Moral Hazard

Principal

Principal: Determines compensation scheme; can only observe outcome (not action)

c1, c2, . . . , cN ß *Compensation*

q1, q2, . . . , qN ß *Outcome*

Action à a1, . . . , aM

Agent: Chooses action a_j; outcome q_i occurs With probability $p(q_i \mid a_j)$

Agent

Moral Hazard (cont…)

Maximize $_{c,a}$ $W(c,a) = \sum_{i=1}^{N} p(q_i \mid a) w(q_i - c_i)$

Subject to $U(c,a) = \sum_{i=1}^{N} p(q_i \mid a) u(c_i, a) \geq U_0$

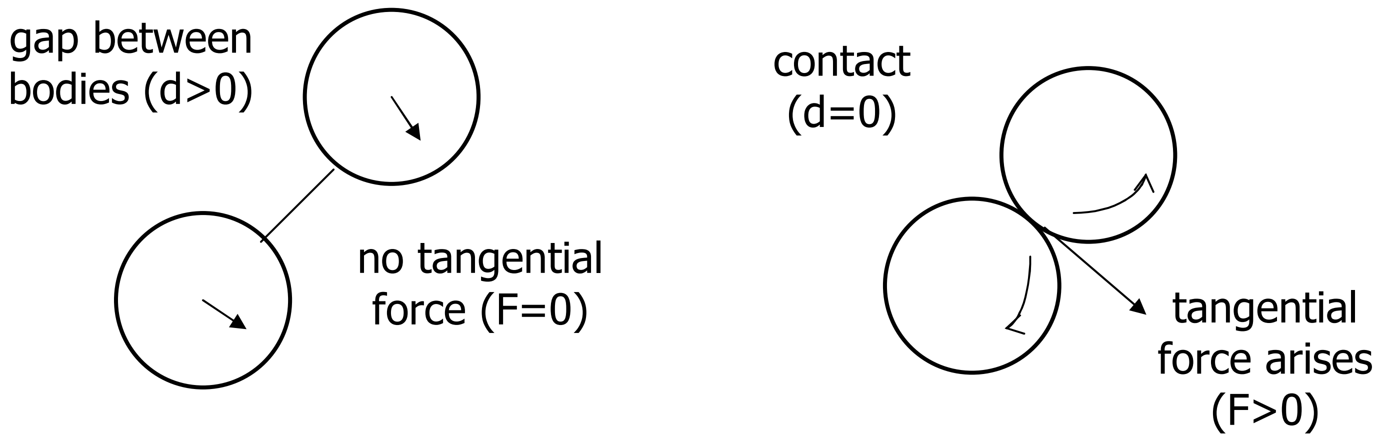
$$a \in \arg\max \left\{ U(c,a) : a \in \{a_1, \ldots, a_M\} \right\}$$

$$c = (c_1, \ldots, c_N) \in \Re_{+}^{N}$$

If a **mixed strategy** profile $(\delta_1,\ldots,\delta_M)$ is introduced for the **agent's action choice** $(a_1,\ldots,a_M)$, this can be **reformulated as an MPEC** by substituting the lower level problem by its optimality conditions (Judd-Su)

Frictional contact

- Tangential force only exists when bodies are in contact



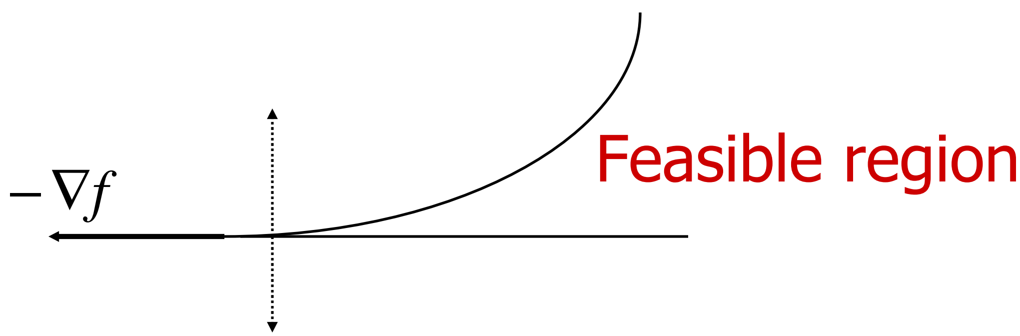
Optimal control (trayectory) of robots with contact

Theoretical/Algorithmic Limits

$$\min f(x) \qquad \nabla f(x) - \nabla g(x)\lambda = 0$$

$$s.t \;\; g(x) \geq 0 \qquad g(x) \geq 0 \;\; \lambda \mathrm{g}(\mathrm{x}) = 0$$

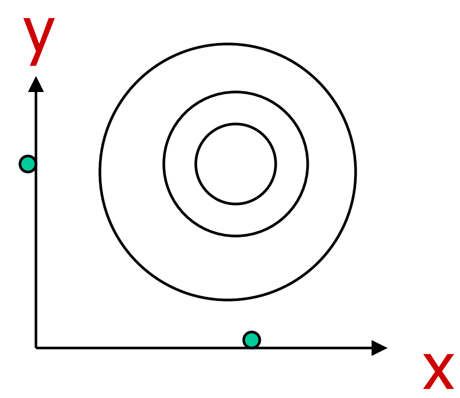
$$\lambda \geq 0$$



- Algorithms based on KKT conditions
- What to do? Perturbation

Equilibrium (Complementarity) Constraints

$$\min \quad (x-1)^2 + (y-1)^2$$
$$x \ge 0, \qquad y \ge 0$$
$$xy = 0$$



(disjunction)

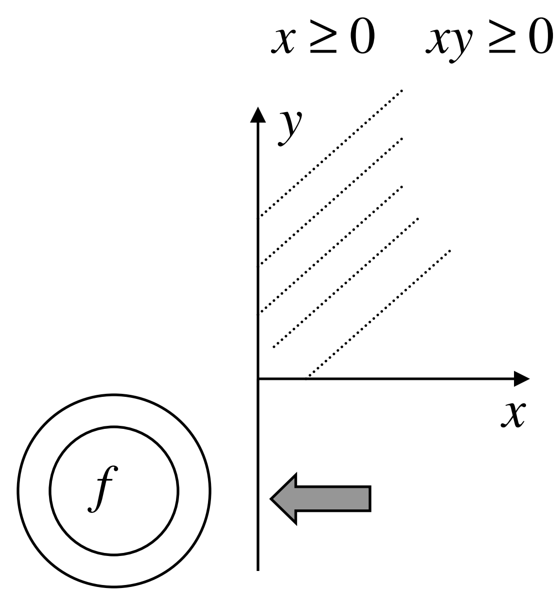
Deficient geometry:

Minimal conditions for

KKT conditions not satisfied

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \begin{pmatrix} y \\ x \end{pmatrix}$$

- Moral Hazard (Judd-Lin)
- Switch-off constraints
- Interior-point methods fail
  - Intrinsic difficulty
- Active-set methods may or may not work



One Solution: L1-Penalty

L1-penalty relaxation of optimization problem with equilibrium constraints

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g(x) \geq 0 \\ & h(x) = 0 \\ & x_1 x_2 = 0 \\ & x_1, x_2 \geq 0 \end{aligned} \quad \longrightarrow \quad \begin{aligned} \min f(x) + \pi\, x_1^T x_2 \\ \text{s.t.} \quad g(x) \geq 0 \\ h(x) = 0 \\ x_1 \geq 0, x_2 \geq 0 \end{aligned}$$

Penalty problem always regular

General technique?

Algorithms and Software

50 years of algorithmic work

Active set

Active set SQP ↓ Active set QP ↓ Linear simplex

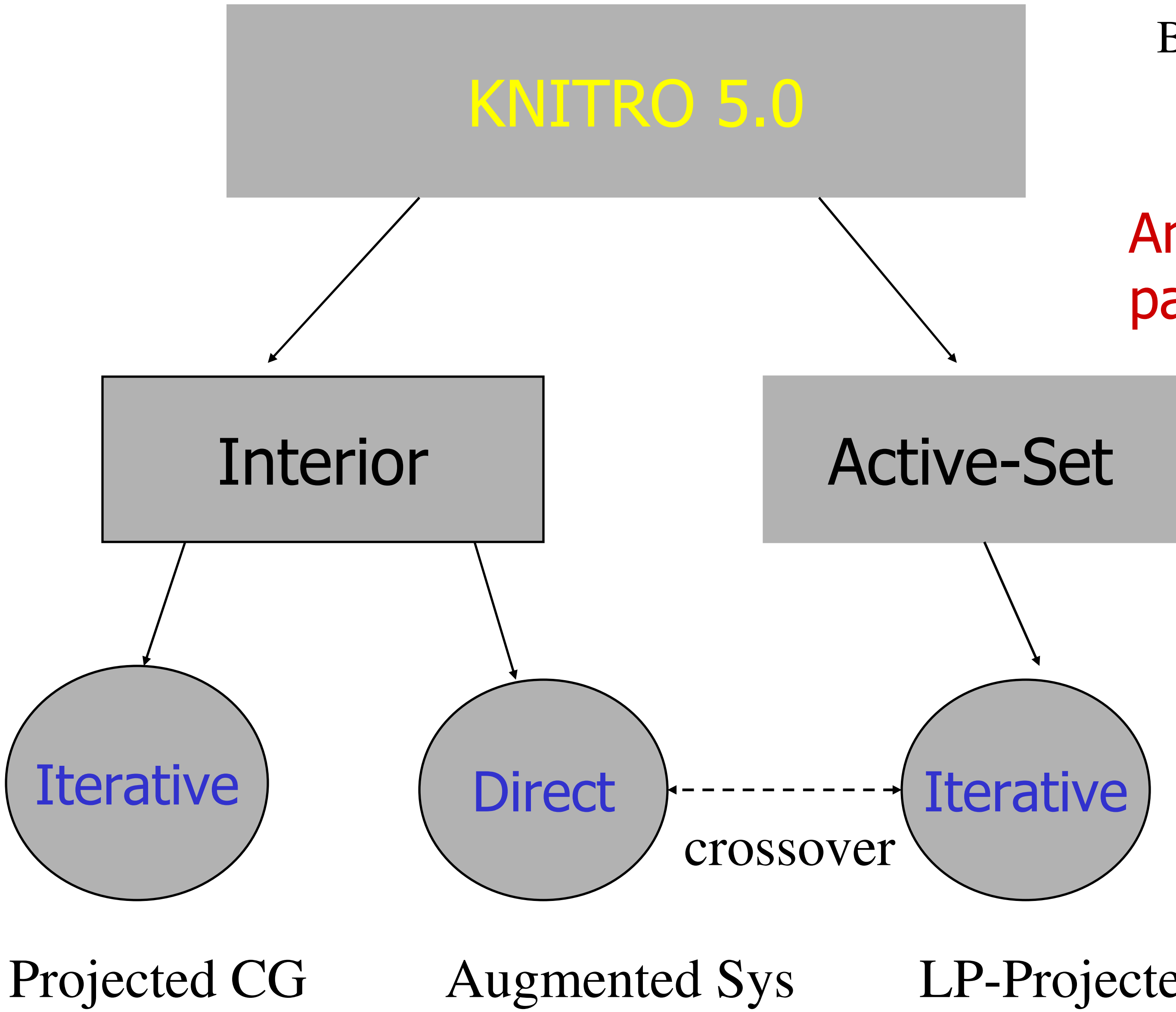
SLQP

Interior (barrier)

Nonlinear ↓ quadratic ↓ Linear

Byrd, Waltz, N.
2006

An integrated package



SLQP

Downward compatibility, Integration

- Applicable and robust for:
  - Unconstrained
  - Bound constrained
  - Equality constrained
  - Nonlinear systems of equations
  - Nonlinear least squares
  - Quadratic programming
  - Linear programming (in progress)
- Multiple derivative options
  - No derivatives
  - First derivatives only
  - First and second (others)

New:
penalty formulation for degenerate constraints

Penalty Methods

- History
- Recent advances

Cutting-edge??

Gould, Orban, Toint 03
Fletcher and Chin 03
Chen and Goldfarb 04
Benson, Vanderbei, Shanno 04
Leyffer, Lopez, N. 04
Anitescu, 2000, 2004
Scholtes 2001
Hu and Ralph, 2002
Benson, Shanno 2005

One Solution: L1-Penalty

L1-penalty relaxation of optimization problem with equilibrium constraints

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g(x) \geq 0 \\ & h(x) = 0 \\ & x_1 x_2 = 0 \\ & x_1, x_2 \geq 0 \end{aligned} \quad \longrightarrow \quad \begin{aligned} \min\ & f(x) + \pi\, x_1^T x_2 \\ \text{s.t.}\ & g(x) \geq 0 \\ & h(x) = 0 \\ & x_1 \geq 0, x_2 \geq 0 \end{aligned}$$

Penalty problem always regular

General technique?

Theoretical Results Algorithms

Soln s of penalty problem à problem solutions
Various stationarity concepts

Scholtes, Ralph, Anitescu, Pang, Luo

Active set methods

Fletcher-Leyffer, Scholtes , Ralph- Zhu, Anitescu

Relaxation (non-penalty):

Friedlander-DeMiguel-Scholtes

Interior-Point methods

Biegler-Ragunathan ,Leyffer-Lopez-N, Vanderbei-Shanno

Classical Unconstrained Formulation

Courant 50s, Fiacco-McCormick 60s

$$\min_x \quad f(x)$$
$$\text{s.t.} \quad h_i(x) = 0,$$
$$g_i(x) \geq 0,$$

$$\boxed{\begin{aligned} &\min_x \ \varphi_\nu(x) = f(x) + \nu \| h(x) \|_2^2 + \nu \| \mathrm{g(x)}^- \|_2^2 \\ &\qquad\qquad g(x)^- = \max\{0, -g(x)\} \end{aligned}}$$

Weaknesses:

- Complicating constraints;
- Infinite family of problems, $\nu \to \infty$
- Ill-conditioning: delicate solution. Need scale invariant method, warm start

Not competitive!

Nonsmooth penalty, L-1

$$\boxed{\varphi_\nu(x) = f(x) + \nu \| h(x) \|_1 + \nu \| g(x)^- \|_1}$$

- One minimization for fixed ν
- Critical points of ϕ are KKT points or infeasible stationary points
- *Almost* parameter free
- Non-smooth, difficult minimization
  - Bundle methods, special techniques?
- Choice of penalty parameter?

Breakthrough: (Fletcher, 1980s)

As in unconstrained minimization:

Create a model of penalty function

Compute steps d by minimizing the model

$$q(d) = \nabla f^T d + \frac{1}{2} d^T W d + \nu \| h + \nabla h^T d \|_1 + \nu \| [g + \nabla g^T d]^- \|_1$$

Non-smooth, but can be reformulated as smooth problem

Note: linear/quadratic -- not quadratic/quadratic

$$q(d) = \nabla f^T d + d^T W d + \nu \| h + \nabla h^T d \|_1 + \nu \| [g + \nabla g^T d]^- \|_1$$

Remove non-smoothness: Linear/quadratic model

Quadratic program

$$\boxed{\begin{array}{ll} \min \nabla f^T d + d^T W d + \nu(u + w + t) & \\ \text{s.t.} \quad h + \nabla h^T d = u - w & \\ \quad g + \nabla g^T d \ge -t, \quad u, w, t \ge 0 & \\ \quad \| d \| \le \Delta \quad \text{(possibly)} & \end{array}}$$

Similar to SQP $\Rightarrow$ SL_1QP *Fletcher*

How is the penalty parameter chosen?

Choose ν_0 and starting point x_0^s

For $k=0,1,\ldots$

Solve penalty problem (linearly cons)

If ||feasibility||< 10^{-6} $\Rightarrow$ Stop

Else

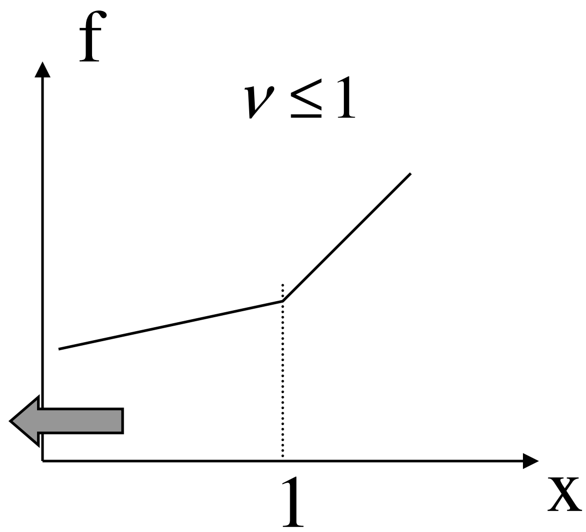
Choose new penalty $\nu_{k+1} > \nu_k$

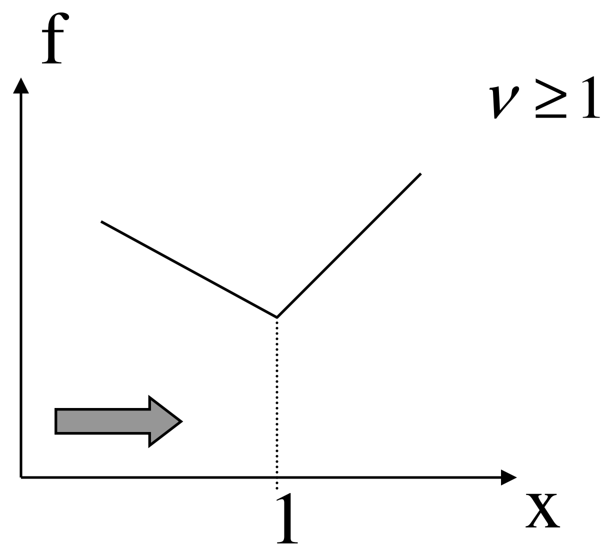
Choose new starting point x_{k+1}^s

End

ADLITTLE LP: Fletcher 1992

- Want ν small to avoid ill-conditioning
- It must be greater than unknown threshold (10^5)
- Hope: if ν is less than threshold, penalty problem infeasible; alerted to increase it
- If ν is about below $10^5/3$:
  - Penalty problem is unbounded
  - Inefficiencies
- Abandon penalty functions $\Rightarrow$ filters



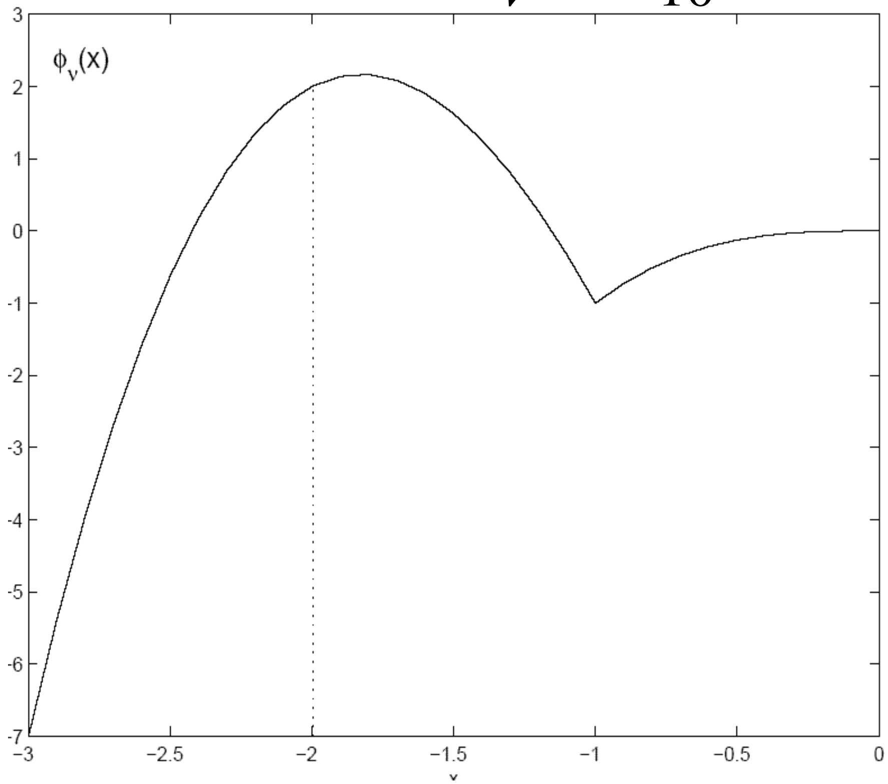


$$\boxed{\min\ x \quad \text{s.t.} \quad x \geq 1} \Rightarrow \min\ x + \nu \max(\,0, 1 - x\,)$$

Can you trust your surrogate….? Consider x=1/2

Only if the improvement in feasibility (to first order) is comparable to the best possible improvement

$\nu \; = \; 10$



$$\boxed{\min x^3 \qquad x \geq -1}$$

x*= - 1 is a local minimizer of ϕ if ν >3.

But ϕ is unbounded as $x \rightarrow -\infty$

For any value of ν there is
A starting point x0 s.t. there is no
Decreasing path from x0
To x*

This example shows that it is not possible to prescribe in advance a value of nu that is adequate t every iteration

- Uconstrained minimization: we control Newton step so that it decreases the objective to first order (descent direction)

$$(\nabla^2 f)\, p = -\nabla f \qquad\qquad \text{pos. def. or trust region}$$

- Similar goal is desirable with respect to the constraints, for constrained optimization…

- Generalize concept to feasibility:
  - Not immediate, requires computation!
  - Implementation in each context
  - It is not a switch but an integral part of the iteration

A Dynamic Strategy
For Selecting the Penalty Parameter

Context: Successive Quadratic Programming Method (Also: Knitro/Active)

QP

$$\min \quad \nabla f^T d + \frac{1}{2} d^T W d$$

$$\text{s.t} \quad \left.\begin{aligned} h + \nabla h^T d &= 0 \\ g + \nabla g^T d &\geq 0 \end{aligned}\right\} \Longrightarrow \text{Incompatible?}$$

$$\| d \|_\infty \leq \Delta$$

Relaxation of quadratic program

QP’

$$\min \nabla f^T d + \nu(u + w + t)$$

$$\text{s.t.} \quad h + \nabla h^T d = u - w$$

Always feasible

$$g + \nabla g^T d \geq -t, \quad u, w, t \geq 0$$

$$\| d \|_\infty \leq \Delta$$

Motivation for new strategy

$$\min \nabla f^T d + \frac{1}{2} d^T W d + \nu(u+w+t)$$
$$\text{s.t.} \quad h + \nabla h^T d = u - w$$
$$g + \nabla g^T d \geq -t, \quad u, w, t \geq 0$$
$$\| d \|_\infty \leq \Delta$$

Idea: if u,w,t
can be zero, do so.
Choose ν accordingly

$$\min \quad (u+w+t)$$
$$\text{s.t.} \quad h + \nabla h^T d = u - w$$
$$g + \nabla g^T d \geq -t,$$
$$u, w, t \geq 0$$
$$\| d \|_\infty \leq \Delta$$

Otherwise solve
Feasibility problem
With nu=infty
A linear program

Adaptive Strategy :

$$
\begin{aligned}
\min \quad & \nabla f^T d + \frac{1}{2} d^T W d + \nu (u + v + t) \\
\text{s.t.} \quad & h + \nabla h^T d = u - v \\
& g + \nabla g^T d \geq -t, \\
& u, v, t \geq 0 \\
& \| d \|_\infty \leq \Delta
\end{aligned}
$$

ν : given

Compute $d(\nu)$

If $u=w=t=0$

  accept ν

Else

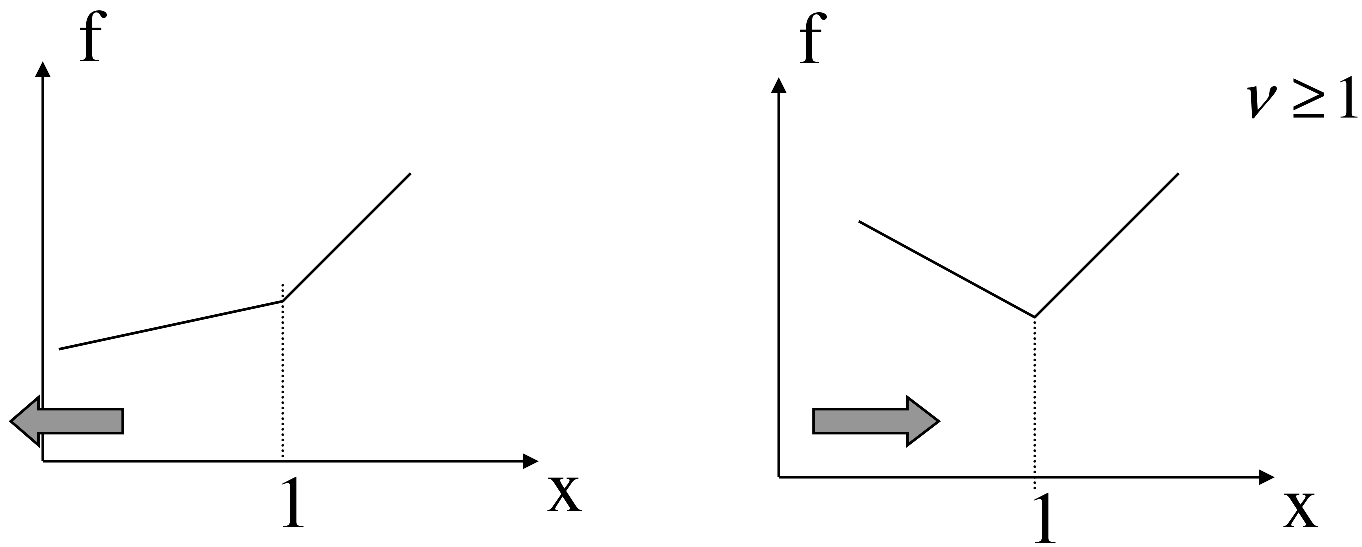
  $\nu = \infty$ compute d^∞

  $m_\infty \equiv \| h + \nabla h^T d^\infty \| + \| [g + \nabla g^T d^\infty]^- \|$

End

Choose ν so that

$$m(0) - m(d^\nu) \geq 0.1[m(0) - m_\infty]$$

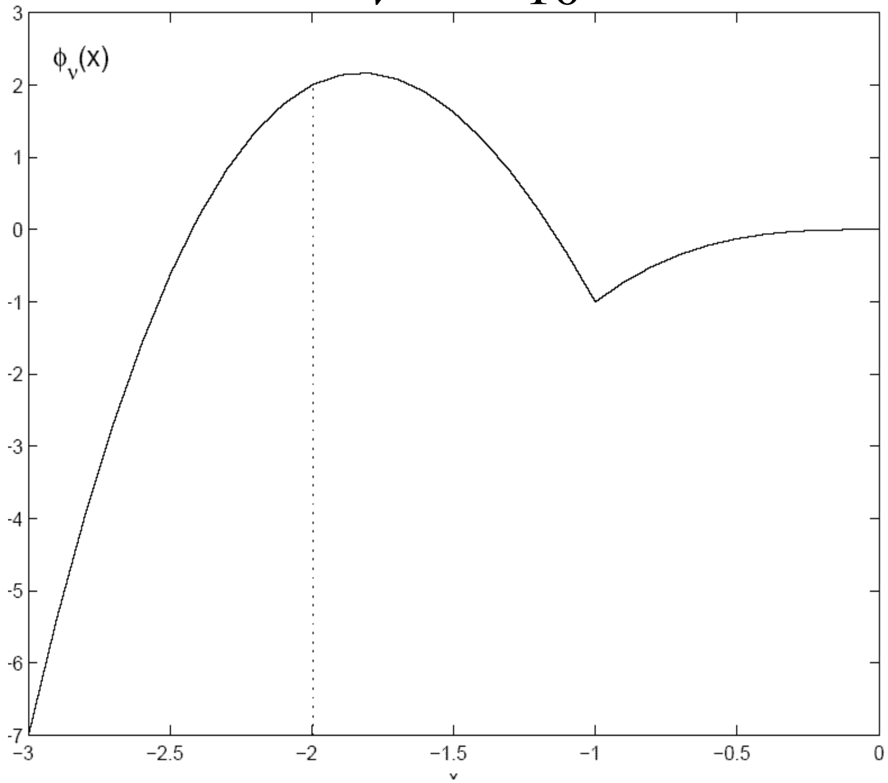


$$\min\ x \qquad \text{s.t.} \quad x \geq 1$$

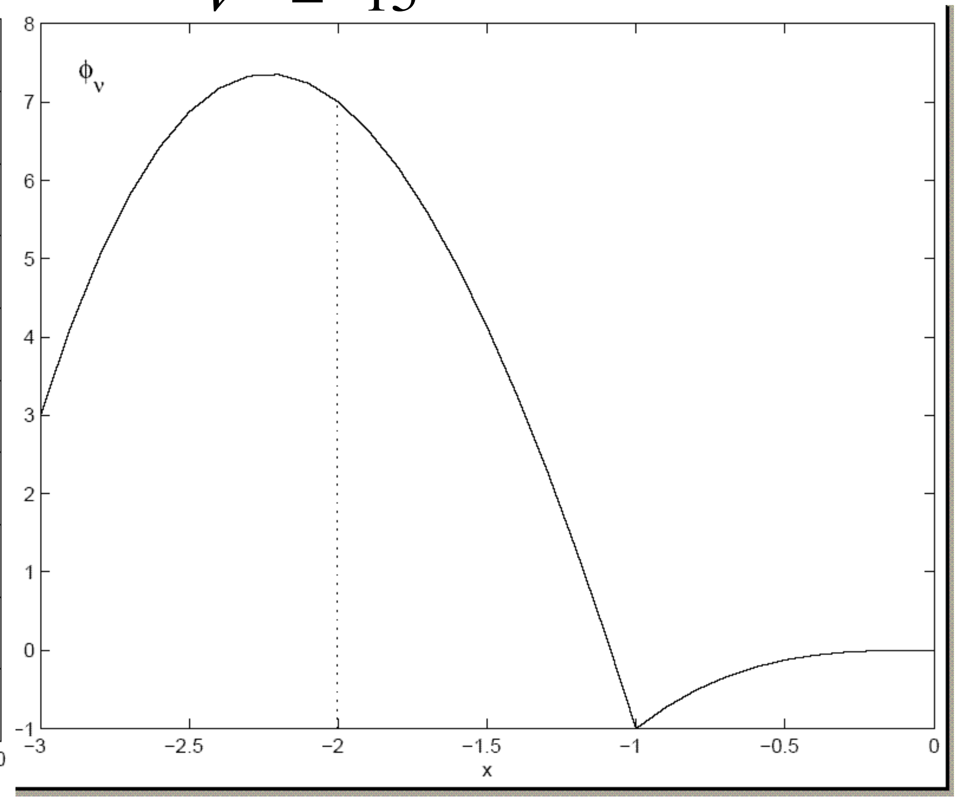
$$m_\infty \equiv \| h + \nabla h^T d^\infty \| + \| [g + \nabla g^T d^\infty]^- \|$$

$$\| d^\infty \| \leq \Delta$$

$\nu = 10$



$\nu = 15$



$$\min x^3 \qquad x \geq -1$$

Our criterion: choose $\nu > 15$

We have expanded
The basin of attraction

Optimality + Feasibility

Improvement in feasibility is not enough. Let

$$l(d) = \nabla f^T d + \nu \| h + \nabla h^T d \|_1 + \nu \| [g + \nabla g^T d]^- \|_1$$

After penalty parameter has been chosen, increase it if necessary s.t.

Promote acceptance of step

$$l(0) - l(d^\nu) \geq 0.1\nu [\, m(0) - m_\infty \,]$$

In trust region notation:

$$pred\ (d) \geq 0.1\nu [cred\ (d)]$$

Knitro-Interior
Dennis, Vicente,
Heinkenschloss,
etc

Adopted in KNITRO/ACTIVE

ADLITTLE LP: Revisited Fletcher 1992

- Running knitro/active:
  - For small trust region radius:
  - Solves in 7 iterations,
  - Penalty from 10 to 10^5 in first 4 iters
  - If trust region includes feasible points, correct adjustment after one LP

Crucial questions:

- Does it actually work in practice?
  - Yes, extensive testing   Waltz
- Possibly solving several subpblems/iteration?
  - Negligible cost
- Can one prove global and local results?
  - Global convergence (Byrd, Gould, Nocedal, Waltz, 2004),
  - active set identification, to be done

The End