New developements in nonlinear programming with perspectives for management sciences

Fabian Bastin Coralia Cartis Cinzia Cirillo Nick Gould Serge Gratton Sven Leyffer Melodie Mouffe Caroline Sainvitu Annick Sartenaer Philippe Toint Dimitri Tomanos Melissa Weber-Mendonça

Department of Mathematics,University of Namur, Belgium

(philippe.toint@fundp.ac.be)

CMS Conference in London, March 2008

Outline

1. Filter methods
   - Constrained optimization
   - Feasibility and least-squares problems
   - Unconstrained and bound-constrained optimization
2. Multilevel algorithms
3. Regularization techniques
4. Non-parametric estimation
5. Conclusions

Outline

1. Filter methods
   - Constrained optimization
   - Feasibility and least-squares problems
   - Unconstrained and bound-constrained optimization
2. Multilevel algorithms
3. Regularization techniques
4. Non-parametric estimation
5. Conclusions

Outline

1. Filter methods
   - Constrained optimization
   - Feasibility and least-squares problems
   - Unconstrained and bound-constrained optimization
2. Multilevel algorithms
3. Regularization techniques
4. Non-parametric estimation
5. Conclusions

Outline

1. Filter methods
   - Constrained optimization
   - Feasibility and least-squares problems
   - Unconstrained and bound-constrained optimization
2. Multilevel algorithms
3. Regularization techniques
4. Non-parametric estimation
5. Conclusions

Outline

1. Filter methods
   - Constrained optimization
   - Feasibility and least-squares problems
   - Unconstrained and bound-constrained optimization
2. Multilevel algorithms
3. Regularization techniques
4. Non-parametric estimation
5. Conclusions

The problem

The general nonlinear programming problem:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & c_{\mathcal{E}}(x) = 0 \\ & c_{\mathcal{I}}(x) \geq 0, \end{array}$$

for $x \in \mathbb{R}^n$, f and c smooth.

Solution algorithms are iterative ($\{x_k\}$) and based on Newton's method (or variant)

very many applications (also in CMS)

Issues: reliability, availability, efficiency

Our objective

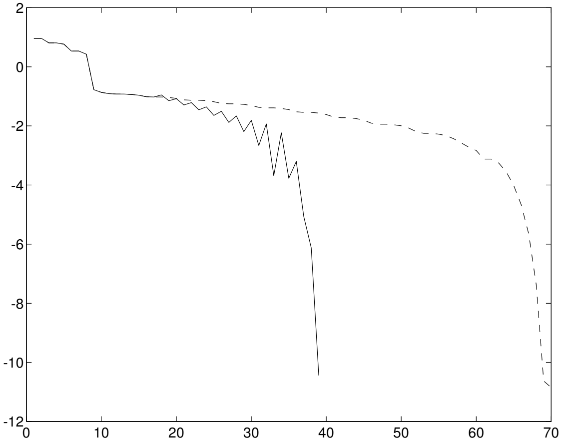
In this talk,

A glimpse into recent (and exciting) new developments

Filter methods

Filter methods

Non-monotonicity



Non-monotonicity often pays!

Introducing the filter

Constrained optimization :

use a good method (SQP) to compute a step s_k from x_k

Ideally

- reduce the objective function $f(x)$
- reduce the constraint violation $\theta(x)$

☺ two potentially conflicting aims ☹

▶ Filter method

Introducing the filter

Constrained optimization :

use a good method (SQP) to compute a step s_k from x_k

Ideally

- reduce the objective function $f(x)$
- reduce the constraint violation $\theta(x)$

☺ two potentially conflicting aims ☹

▶ Filter method

Introducing the filter

Constrained optimization :

use a good method (SQP) to compute a step s_k from x_k

Ideally

- reduce the objective function $f(x)$
- reduce the constraint violation $\theta(x)$

☺ two potentially conflicting aims ☹

▶ Filter method

Accepting a new iterate

Idea of Fletcher and Leyffer

Replace the question

What is a better point ?

by

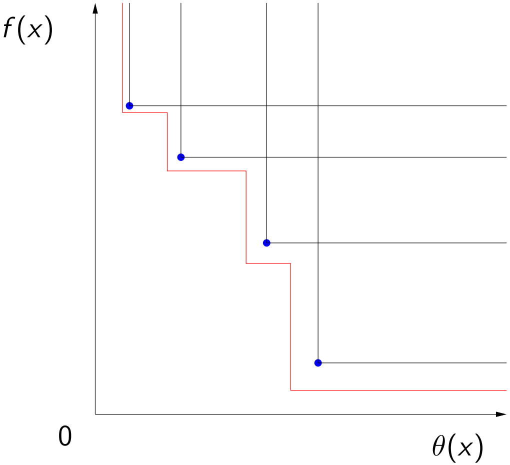
What is a worse point ?

Of course, y is "worse" than x if it is dominated by x, i.e., when

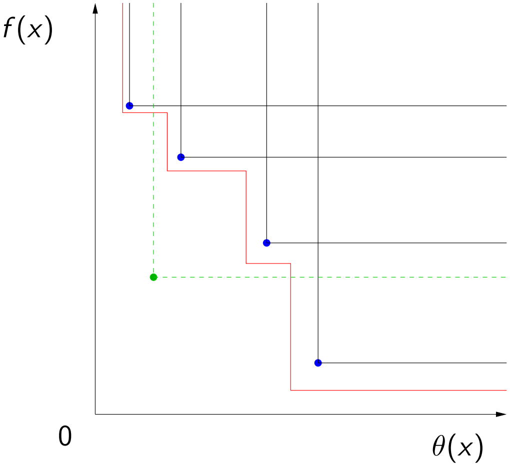
$$f(x) \leq f(y) \quad \text{and} \quad \theta(x) \leq \theta(y)$$

Accept or reject a trial point ?

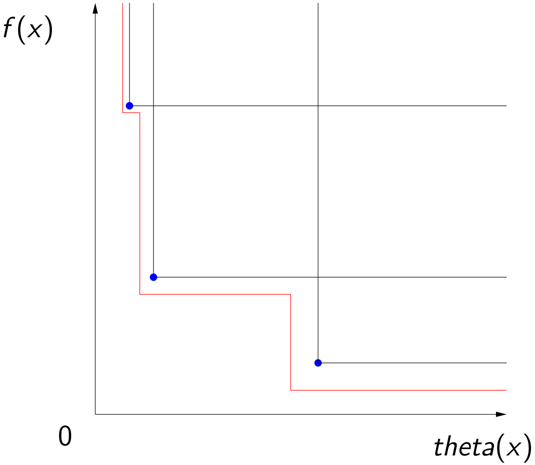
View of a standard filter



View of a standard filter



View of a standard filter



Handling many objectives

Gould, Leyffer and T.

Feasibility

Find x such that

$$c_{\mathcal{E}}(x) = 0 \quad \text{and} \quad c_{\mathcal{I}}(x) \geq 0$$

Least-squares

Find x such that

$$\min_{x \in \mathbf{R}^n} \sum_{i \in \mathcal{E} \cup \mathcal{I}} \theta_i(x)^2$$

▶ More dimensions in the filter space . . .

Handling many objectives

Gould, Leyffer and T.

Feasibility

Find x such that

$$c_{\mathcal{E}}(x) = 0 \quad \text{and} \quad c_{\mathcal{I}}(x) \geq 0$$

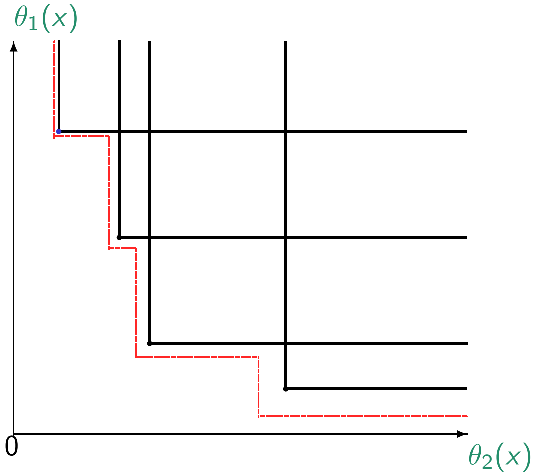
Least-squares

Find x such that

$$\min_{x \in \mathbf{R}^n} \sum_{i \in \mathcal{E} \cup \mathcal{I}} \theta_i(x)^2$$

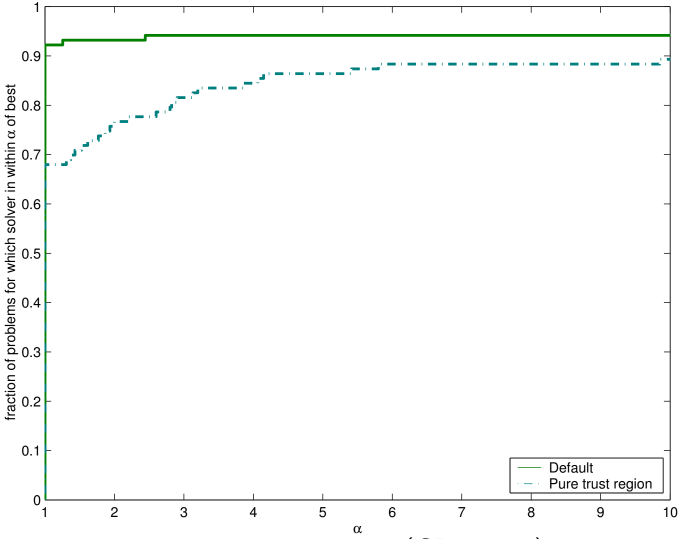
▶ More dimensions in the filter space . . .

The obvious picture



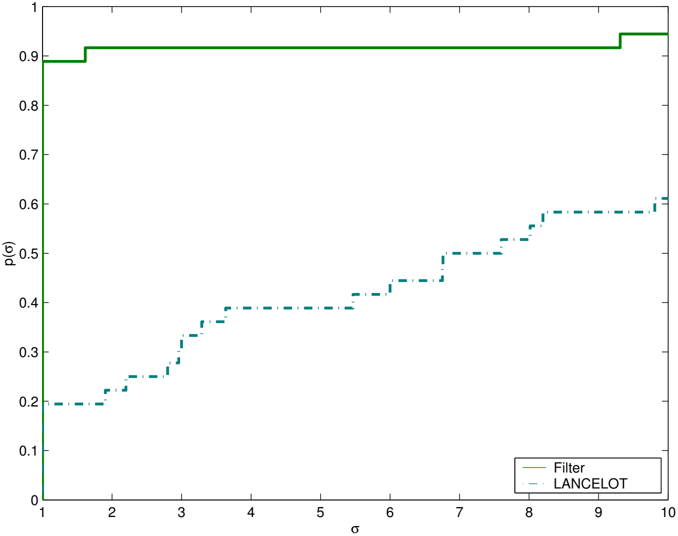
(full dimension vs. grouping)

Numerical experience (1)



Filter vs. trust-region (CPU time)

Numerical experience (2)



Filter vs. LANCELOT B (CPU time)

The other extreme case

Gould, Sainvitu and T.

Unconstrained optimization

$$\min_{x \in \mathbf{R}^n} f(x)$$

Simple-bound constrained optimization

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & l \leq x \leq u \end{array}$$

Combined methods:

Trust-region + filter + projected gradient

▶ Multidimensional filter

A gradient multidimensional filter

Accept x_k^+ more often

A point x dominates a point y whenever

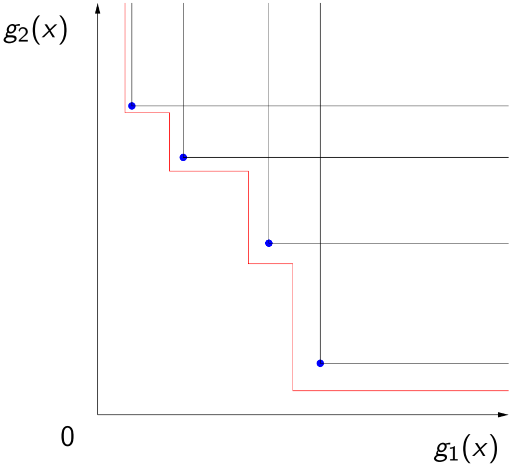
$$|g_i(x)| \leq |g_i(y)| \quad \forall\, i = 1, \ldots, n$$

Remember non-dominated points $\Rightarrow$ FILTER

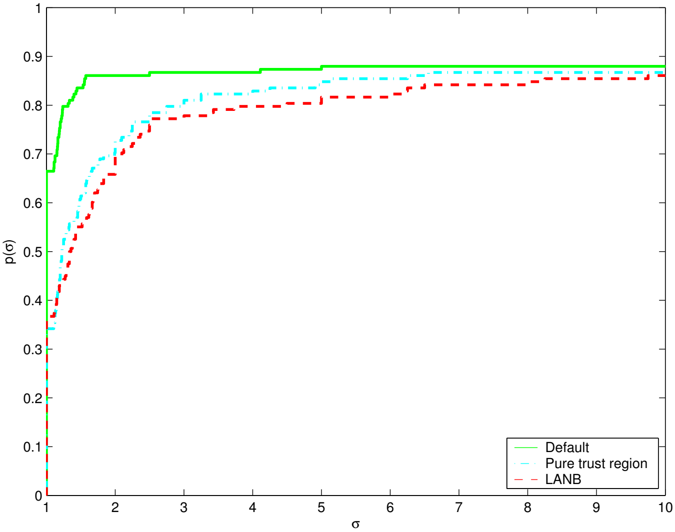
Accept a new iterate ?

if it is not dominated by any other iterate in the filter

Haven't we seen this before?

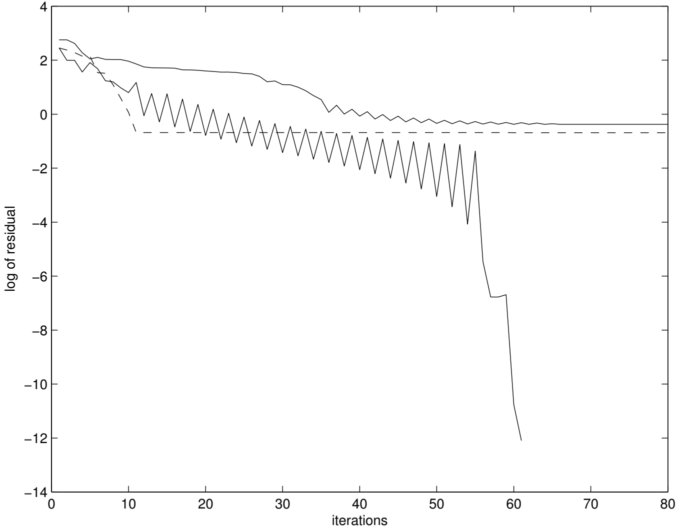


Numerical experience (unconstrained)



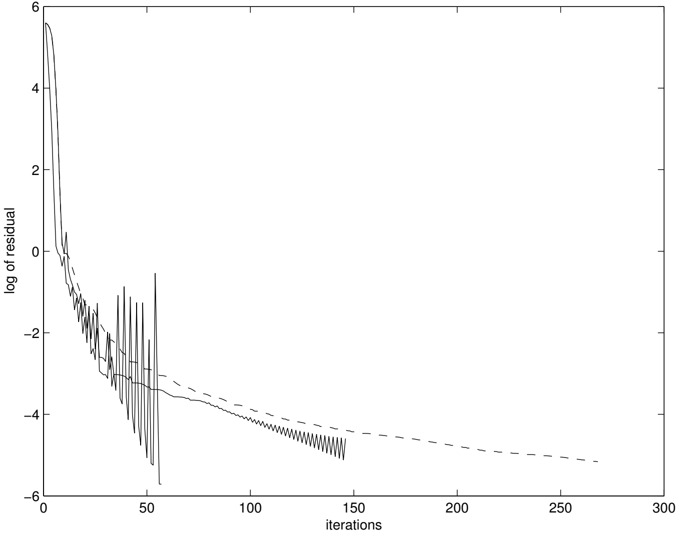
Filter vs. trust-region and LANCELOT B (iterations)

Numerical experience: HEART6



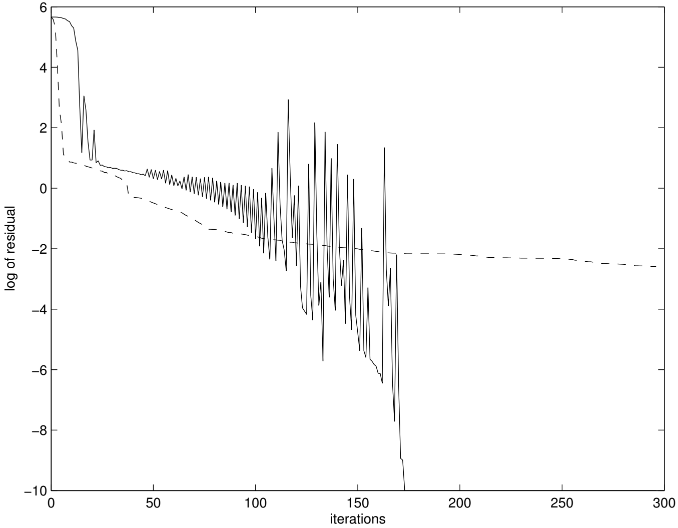
Filter vs. trust-region and LANCELOT B

Numerical experience: EXTROSNB



Filter vs. trust-region and LANCELOT B

Numerical experience: LOBSTERZ



Filter vs. trust-region

Multilevel algorithms

Multilevel algorithms

Motivation

- Optimization of continuous problems occurs in a many applications: shape optimization, data assimilation, control problems, . . .
- Recent optimization methods have been designed to cope with these problems, including multilevel/multigrid algorithms.
- These algorithms involve the computation of a hierarchy of problem descriptions, linked by known operators.

Here: focus on unconstrained/ bound-constrained optimization:

$$\min_{(x \geq 0)} f(x)$$

Hierarchy of problem descriptions

Can we use a structure of the form:

Finest problem description

Restriction $\downarrow$ R — P $\uparrow$ Prolongation

Fine problem description

Restriction $\downarrow$ R — P $\uparrow$ Prolongation

$\cdots$

Restriction $\downarrow$ R — P $\uparrow$ Prolongation

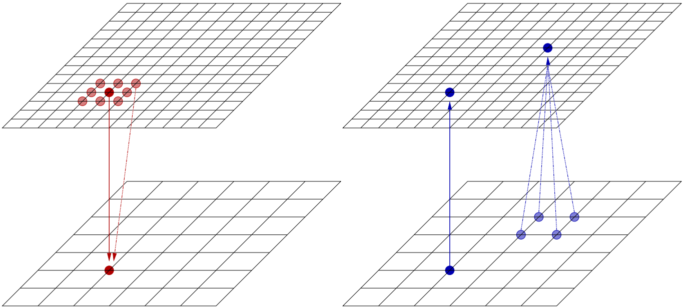
Coarse problem description

Restriction $\downarrow$ R — P $\uparrow$ Prolongation

Coarsest problem description

Grid transfer operators

$$R_i : \quad \mathbb{R}^{n_i} \quad \longrightarrow \quad \mathbb{R}^{n_{i-1}} \qquad \text{Restriction}$$
$$P_i : \quad \mathbb{R}^{n_{i-1}} \quad \longrightarrow \quad \mathbb{R}^{n_i} \qquad \text{Prolongation}$$

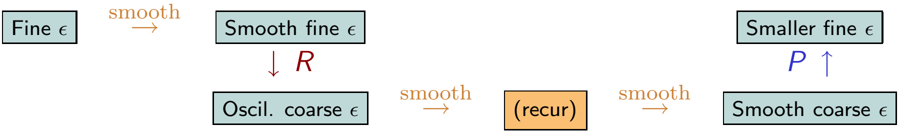


Three keys to multigrid algorithms

- oscillatory components of the error are representable on fine grids, but not on coarse grids
- iterative methods reduce oscillatory components much faster than smooth ones
- smooth on fine grids $\longrightarrow$ oscillatory on coarse ones

How to exploit these keys

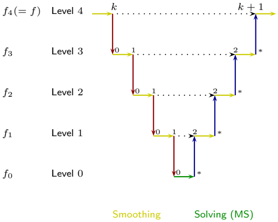
Annihilate oscillatory error level by level:

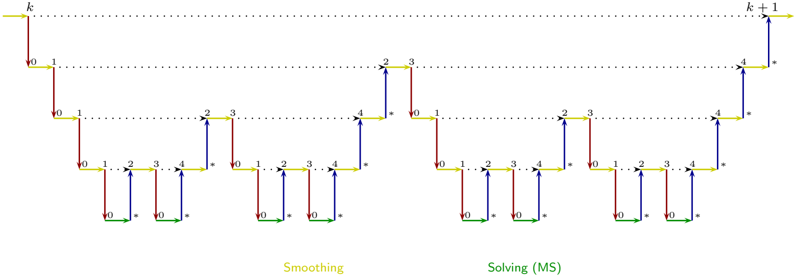


Note: P and R are not othogonal projectors!

A very efficient method for some linear systems
(when A(smooth modes) $\in$ smooth modes)

Iteration structure across levels

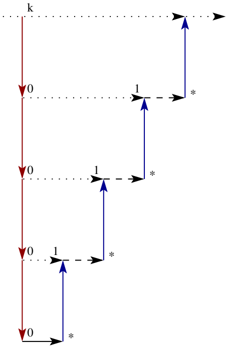




Mesh refinement, as different from. . .

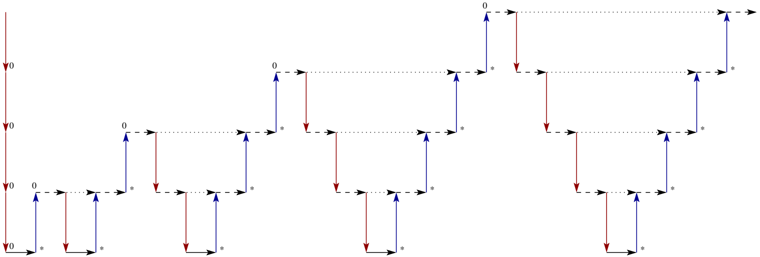
Computing good starting points:

- Solve the problem on the coarsest level
  $\Rightarrow$ Good starting point for the next fine level
- Do the same on each level
  $\Rightarrow$ Good starting point for the finest level
- Finally solve the problem on the finest level



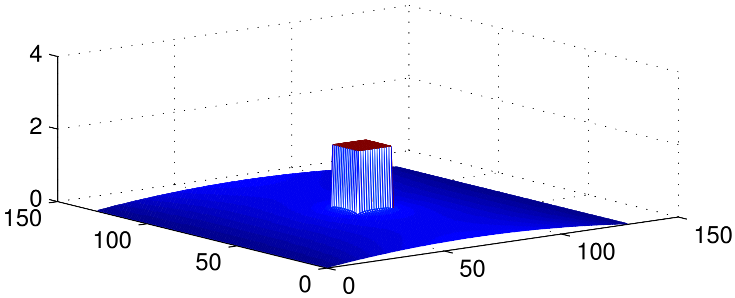
. . . V-cycles and Full Multigrid (FMG)

- FMG : Combination of mesh refinement and V-cycles

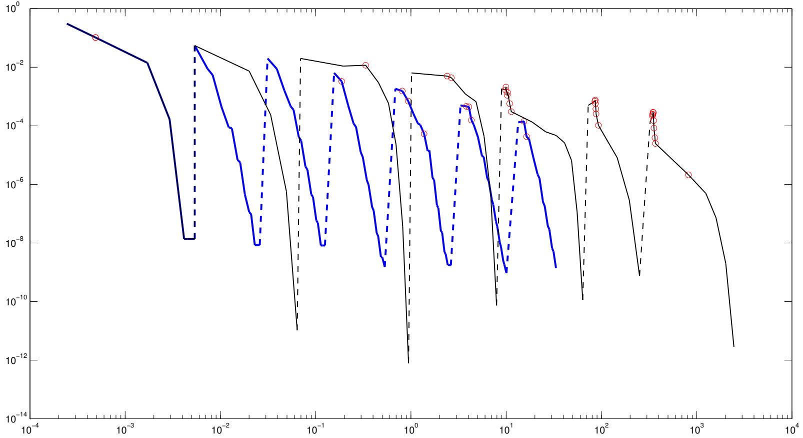


Results on an obstacle problem ($n = 127^2$, 6 levels)

	unconstrained			bound-constrained	
	Mesh ref.	$RMTR_2$	$RMTR_\infty$	Mesh ref.	$RMTR_\infty$
nit	1057	23	10	2768	214
nf	23	38	15	649	240
ng	16	28	14	640	236
nH	17	20	6	32	101



Equivalent iterations on the minimum surface problem



mesh-refinement FMG

Regularization techniques

Regularization techniques

Unconstrained optimization — a "mature" area?

$$\underset{x \in \mathbf{R}^n}{\text{minimize}} \quad f(x) \quad \text{where} \quad f \in C^1 \quad (\text{maybe} \quad C^2 \;)$$

Currently two main competing (but similar) methodologies

- **Linesearch methods**
  - compute a descent direction s_k from x_k
  - set $x_{k+1} = x_k + \alpha_k s_k$ to improve f
- **Trust-region methods**
  - compute a step s_k from x_k to improve a model m_k of f within the trust-region $\|s\| \leq \Delta$
  - set $x_{k+1} = x_k + s_k$ if m_k and f "agree" at $x_k + s_k$
  - otherwise set $x_{k+1} = x_k$ and reduce the radius Δ

Is there anything more to say?

Recently, Nesterov and Polyak (2006) observed the following: if

- f has gradient g and globally Lipschitz continuous Hessian H with constant $2L$

Taylor, Cauchy-Schwarz and Lipschitz imply

$$\begin{aligned} f(x+s) &= f(x) + \langle s, g(x)\rangle + \tfrac{1}{2}\langle s, H(x)s\rangle \\ &\quad + \int_0^1 (1-\alpha)\langle s, [H(x+\alpha s) - H(x)]s\rangle \, d\alpha \\ &\leq \underbrace{f(x) + \langle s, g(x)\rangle + \tfrac{1}{2}\langle s, H(x)s\rangle + \tfrac{1}{3}L\|s\|_2^3}_{m(s)} \end{aligned}$$

$\Longrightarrow$ reducing m from $s=0$ improves f since $m(0)=f(x)$.

The cubic regularization

Change from

$$\min_s \quad f(x) + \langle s, g(x)\rangle + \tfrac{1}{2}\langle s, H(x)s\rangle \;\; \text{s.t.} \;\; \|s\| \le \Delta$$

to

$$\min_s \quad f(x) + \langle s, g(x)\rangle + \tfrac{1}{2}\langle s, H(x)s\rangle + \tfrac{1}{3}\sigma\|s\|^3$$

σ is the (adaptive) regularization parameter

The main features of adaptive cubic regularization

And the result is. . .

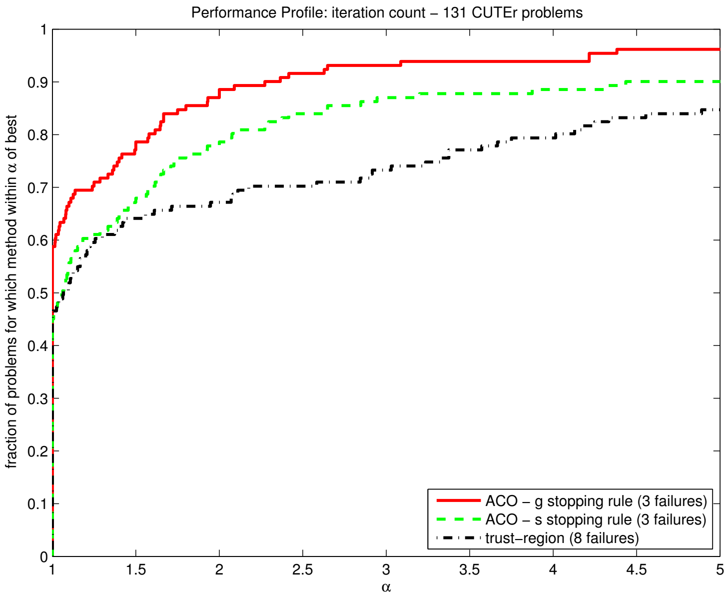
longer steps on ill-conditioned problems

similar (very satisfactory) convergence analysis

best known worst-case complexity for nonconvex problems

excellent performance and reliability

Numerical experience — small problems using Matlab



Non-parametric estimation

Non-parametric estimation

The context and problem

Agent decision modelling

Random utility models

Max likelihood parameter estimation

- Problem: Value of parameter(s) sometimes unrealistic!
  Example: negative VOT. . .
- symmetric distribution assumed in random parts of utility!

Using advanced optimization tools. . .

What can we do?

- avoid irrealistic distribution assumptions
- estimate non-parametric distributions from the data
- need for an efficient representation of non-parametric distributions

cubic spline representation of inverse CDF

$\Rightarrow$ order constraint on the knot values in likelihood maximization!

efficient projection based trust-region algorithm

And the result is. . .

Excellent estimation of asymmetric distribution

A nice application on the behaviour of the BoJ of the FX market

▶ See Fabian Bastin's talk!◀

Thus . . .

Good use of advanced optimization = progress in CMS

Conclusions

- a vibrant field with new perspectives
- expected: more reliable and more efficient methods
- ... available to a larger public
- ... and applications in more challenging domains
- CMS continues to be the source of
  - interesting applications
  - methodologicaal inspiration

Thank you for your attention !

(see `http://perso.fundp.ac.be/~phtoint/publications.html` for references)