Filter methods: a tribute to Roger Fletcher

Philippe Toint (with S. Leyffer, N. Gould and C. Sainvitu)

Namur Center for Complex Systems (naXys), University of Namur, Belgium

(philippe.toint@unamur.be)

Beijing, August 2016

This talk is a tribute to Roger Fletcher, a friend and a lasting inspiration.

Some of his ideas (my biased and partial view)

- the Fletcher-Reeves conjugate gradients
- Fletcher-Matthews stable LU factorization updating
- BFGS
- sequential ℓ_1 quadratic programming $S\ell_1QP$
- restricted step (trust-regions) methods
- "least-squares" differentiable penalty function
- filter methods (with Sven)
- ...

+ a wonderful textbook

Nonlinear optimization

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 ($\{x_k\}$)
- based on Newton's method (or variant)

$\Rightarrow$ global convergence issues

Monotonicity (1)

Global convergence theoretically ensured by

- some global measure ...
  - unconstrained : $f(x_k)$
  - constrained : merit function at x_k
- ... with strong monotonic behaviour

(Lyapunov function)

Also practically enforced by

- algorithmic safeguards around Newton method (linesearches, trust regions)

Monotonicity (2)

But

classical safeguards limit efficiency!

Question:

design less obstructive safeguards

while

- ensuring better numerical performance
  (the Newton Liberation Front !)
- continuing to guarantee
  global convergence properties

Non-monotone methods

Typically:

- abandon strict monotonicity of usual measures
- but insist on average behaviour

linesearch:

- Chamberlain, Powell, Lemarechal, Pedersen (1982)
- Grippo, Lampariello, Lucidi, Facchinei (1986, 1989, 1991, 1992,. . .)
- Panier, Tits, Bonnans, Zhou (1991, 1992), T. (1996), . . .

trust region:

- Deng, Xiao, Zhou (1992, 1993, 1994, 1995)
- T. (1994, 1997), Conn, Gould, T. (2000)
- Ke, Han, Liu (1995, 1996), Burke, Weigmann (1997), Yuan (1999), . . .

Non-monotone trust-regions

Idea:

$$f(x_{k+1}) < f(x_k) \text{ replaced by } f(x_{k+1}) < f_{r(k)}$$

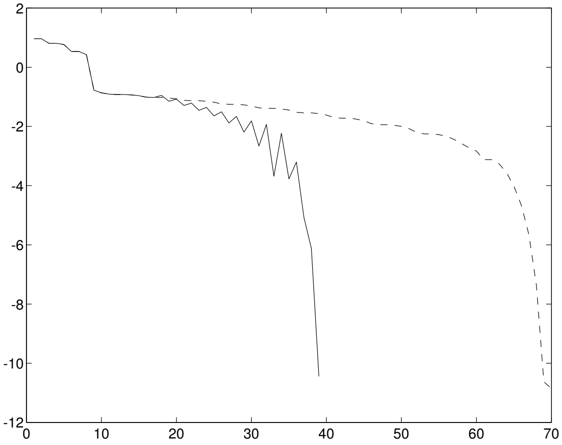
with

$$f_{r(k)} < f_{r(k-1)}$$

Further issues:

- suitably define $r(k)$
- adapt the trust-region algorithm:
  also compare achieved and predicted reductions since reference iteration

An unconstrained example



Monotone and non-monotone TR (LANCELOT B)

Introducing the filter

A fruitful alternative: **filter methods**

Constrained optimization :

using the SQP step, at the same time:

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

$\Rightarrow$ CONFLICT

The filter point of view

Fletcher and Leyffer replace question:

What is a better point?

by:

What is a worse point?

Of course, y is worse than x when

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

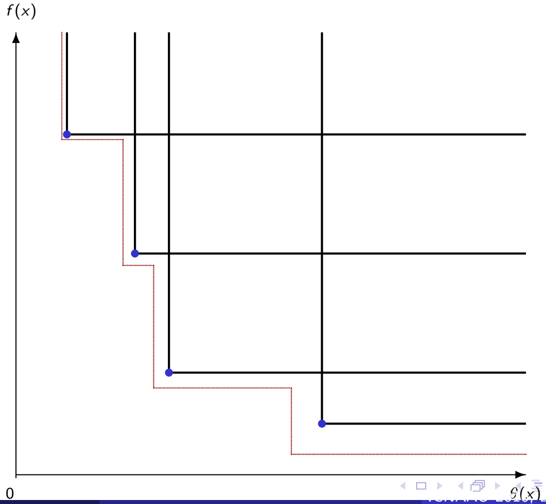
(y is dominated by x)

When is $x_k + s_k$ acceptable?

The standard filter

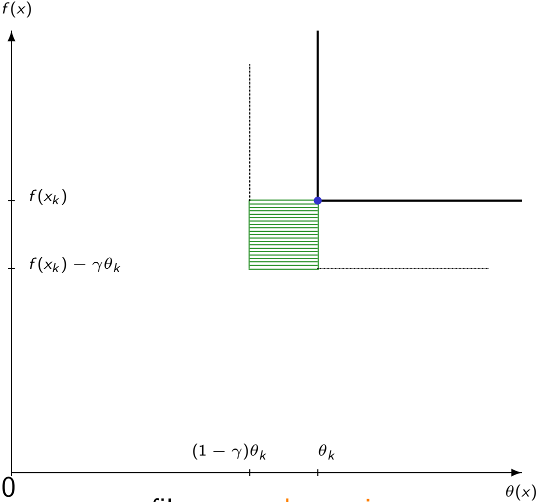
Idea: accept non-dominated points

no monotonicity of merit function implied



Filling up the standard filter

Note: filter area is bounded in the (f, θ) space!



$\Rightarrow$ filter area decreasing

The unconstrained feasibility problem

Feasibility

Find x such that

$$c(x) \geq 0$$

$$e(x) = 0$$

for general smooth c and e.

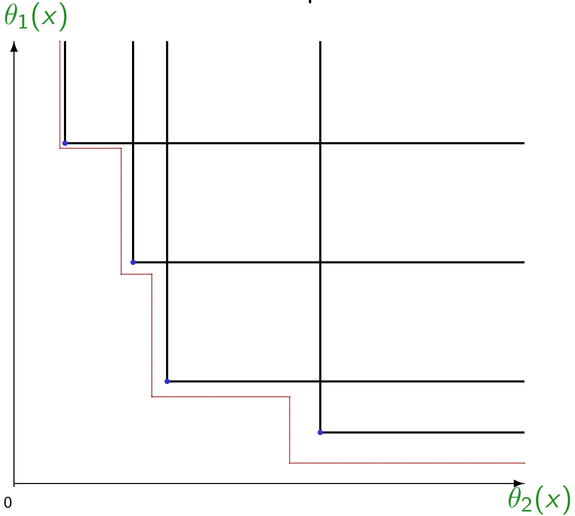
Least-squares

Find x such that

$$\min \sum \theta_i^2$$

A multidimensional filter (1)

(Simple) idea: more dimensions in filter space



(full dimension vs. grouping)

A multidimensional filter (2)

Additionally

- possibly consider unsigned filter entries
- use TR algorithm when
  - trial point unacceptable
  - convergence to non-zero solution

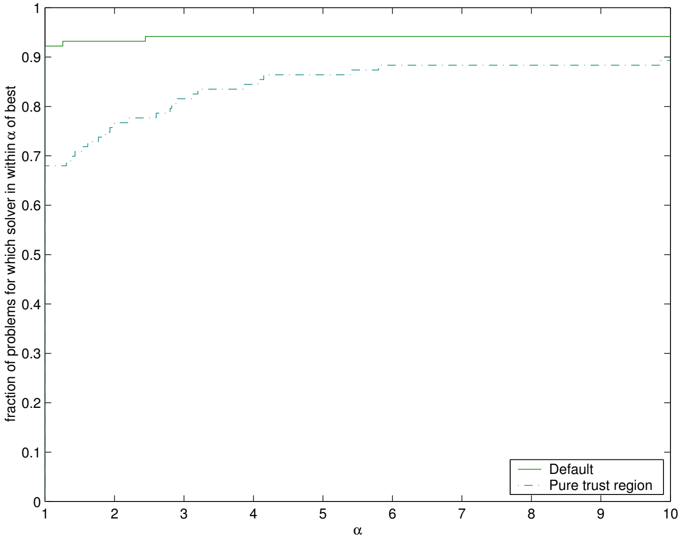
($\Rightarrow$ "internal" restoration)

sound convergence theory

Numerical experience: FILTRANE

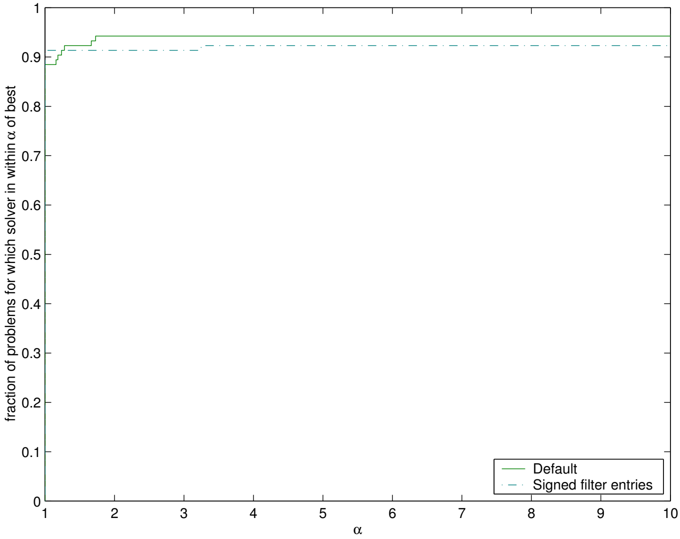
- Fortran 95 package
- large scale problems (CUTEr interface)
- includes several variants of the method
  - signed/unsigned filters
  - Gauss-Newton, Newton or adaptive models
  - pure trust-region option
  - uses preconditioned conjugate-gradients + Lanczos for subproblem solution
- part of the GALAHAD library

Numerical experience (1)



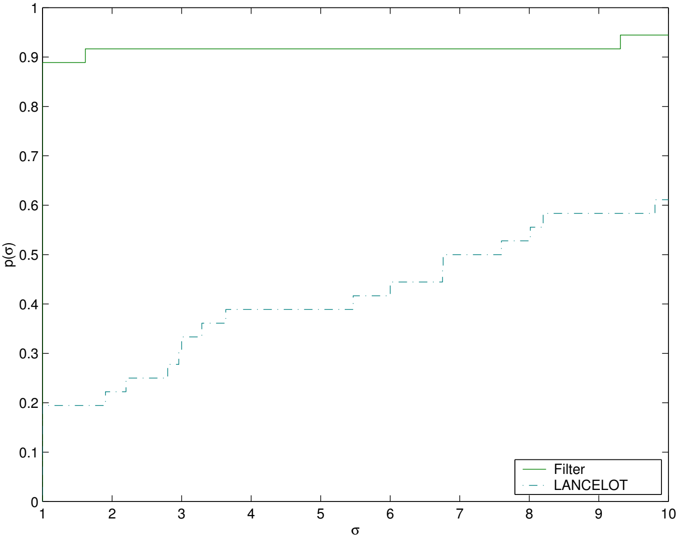
Filter vs. trust-region (CPU time)

Numerical experience (2)



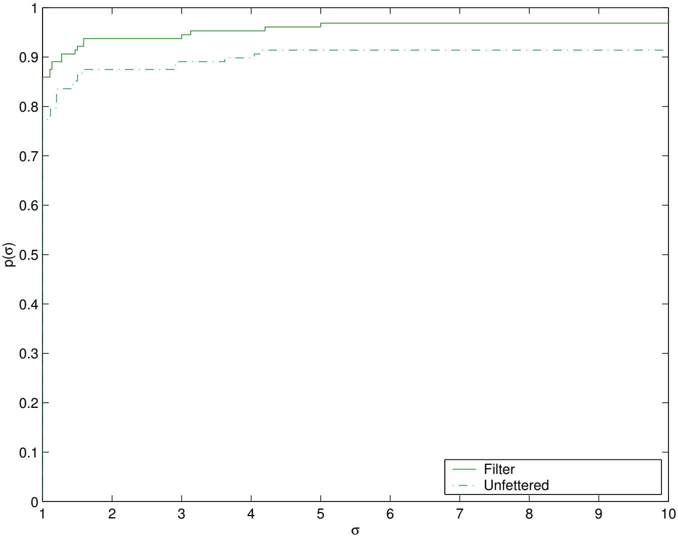
Allowing unsigned filter entries (CPU time)

Numerical experience (3)



Filter vs. LANCELOT B (CPU time)

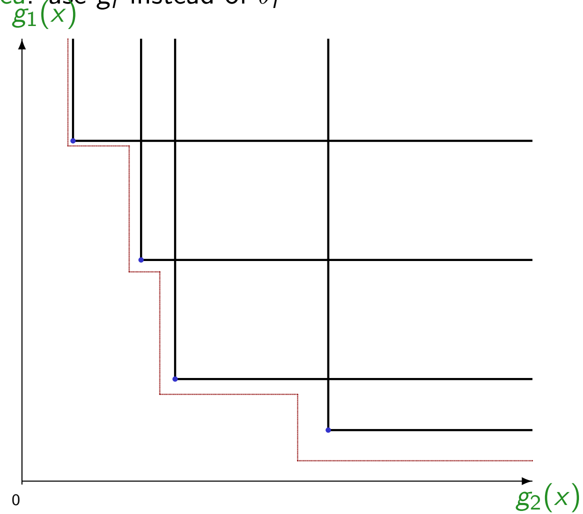
Numerical experience (4)



Filter vs. free Newton (CPU time)

Filter for unconstrained opt.

Again simple idea: use g_i instead of θ_i



(full dimension vs. grouping)

A few complications. . .

But . . .

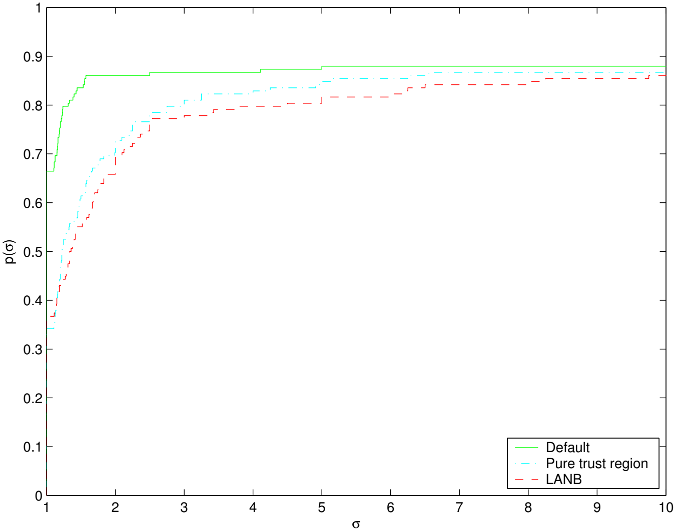
$g(x) = 0$ not sufficient for nonconvex problems!

When negative curvature found:

- reset filter
- set upper bound on acceptable $f(x)$

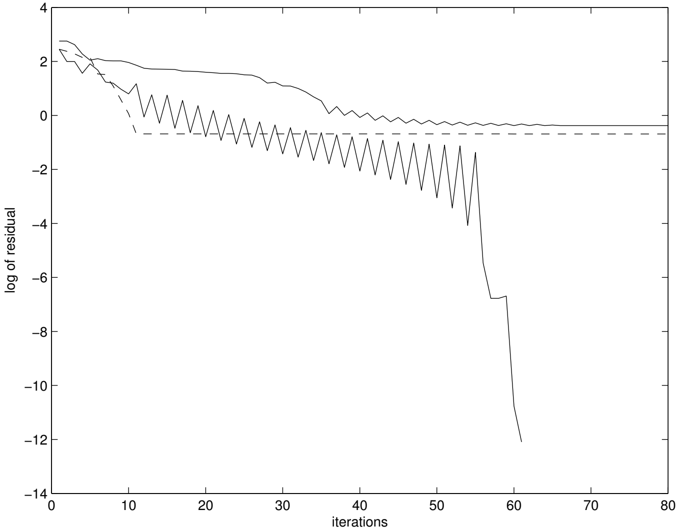
reasonable convergence theory

Numerical experience (1)



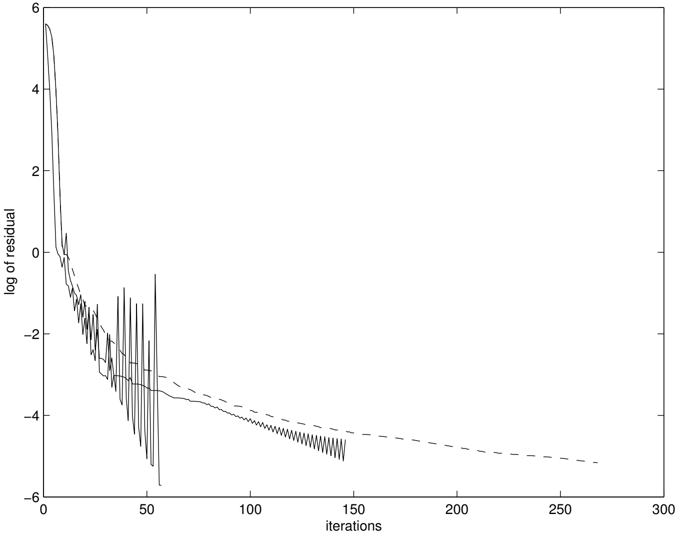
Filter vs. trust-region and LANCELOT B (iterations)

Numerical experience: HEARTS



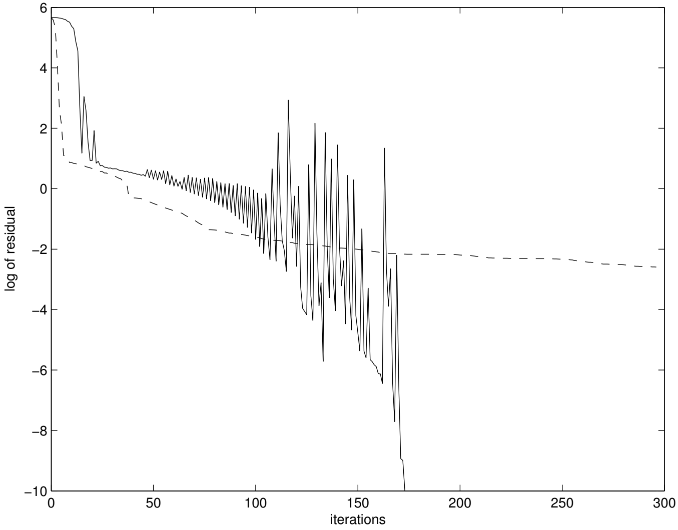
Filter vs. trust-region and LANCELOT B

Numerical experience: EXTROSNB



Filter vs. trust-region and LANCELOT B

Numerical experience: LOBSTERZ



Filter vs. trust-region

Conclusions

non-monotonicity definitely helpful

Newton's behaviour unexplained

Thanks to Roger (and Sven) for the filter idea

Thank you for your attention