The filter idea and its application to nonlinear equations, nonlinear inequalities and nonlinear least-squares

Philippe Toint
(University of Namur, Belgium)
(with N. Gould)

Bath, January 2003

Outline (1)

- Filter SQP for nonlinear programming
  - the problem
  - from merit functions to dominance
  - the filter
  - the SQP filter algorithm
  - the experience so far

Outline (2)

- Filter for the feasibility problem
  - the problem
  - mutidimensional filters
  - the algorithm
  - a few words about convergence theory
  - some numerical results

- Conclusions

The NLP problem

The general nonlinear programming problem

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & c(x) \geq 0 \end{array}$$

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

(insist, for now, on inequality constraints)

Define the maximum constraint violation :

$$\theta(x) = \max[0, -c_1(x), \ldots, -c_m(x)].$$

Classically: merit function

$$\phi(x, \mu) = f(x) + \frac{1}{\mu}\theta(x)$$

(or similar)

Problems with merit functions

- the coherence between strategy to define a step and the merit function Φ (Maratos effect)

- the practical rules to manage the penalty parameter μ
  - initial value
  - (non-)monotone updating rules

$\Rightarrow$ Can we do without a merit function?

Conflicting objectives and dominance

At the same time:

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

$\Rightarrow$ conflict

Classical question: what is a better point?

Rephrase as: What is a worse point?

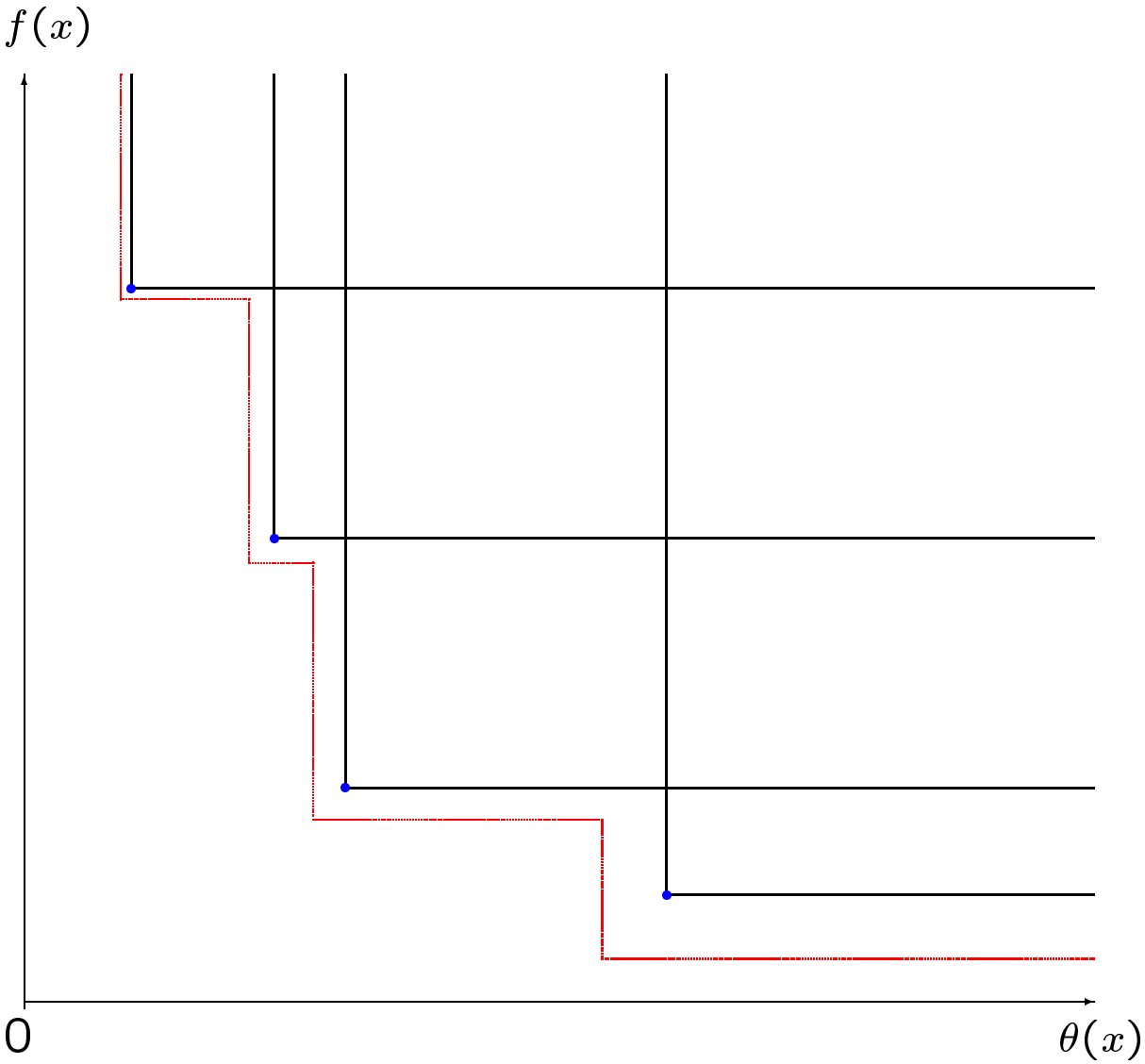
$$f(x) \leq f(y) \text{ and } \theta(x) \leq \theta(y)$$

(y dominated by x)

The filter algorithm

Idea: Build a numerical method where non-dominated points are accepted.

remember accepted points $\Rightarrow$ filter ($\mathcal{F}$)



The filter algorithm (2)

Step selection:

- (approximate) SQP step $s_k = n_k + t_k$
  - satisfies linearized constraints (when possible) $\rightarrow n_k$ (normal step)
  - reduces a model m_k of the Lagrangian $\rightarrow t_k$ (tangential step)
- trust-region to ensure global convergence (Δ_k)

TRQP(x_k, Δ_k):

$$\begin{array}{ll} \text{minimize} & m_k(s) \\ \text{subject to} & c_k + A_k s \geq 0 \\ \text{and} & \|s\| \leq \Delta_k, \end{array}$$

The filter algorithm (3)

What if TRQP(x_k, Δ_k) is incompatible?

- forget the objective function

- reduce $\theta(x)$
  (typically using another TR method)

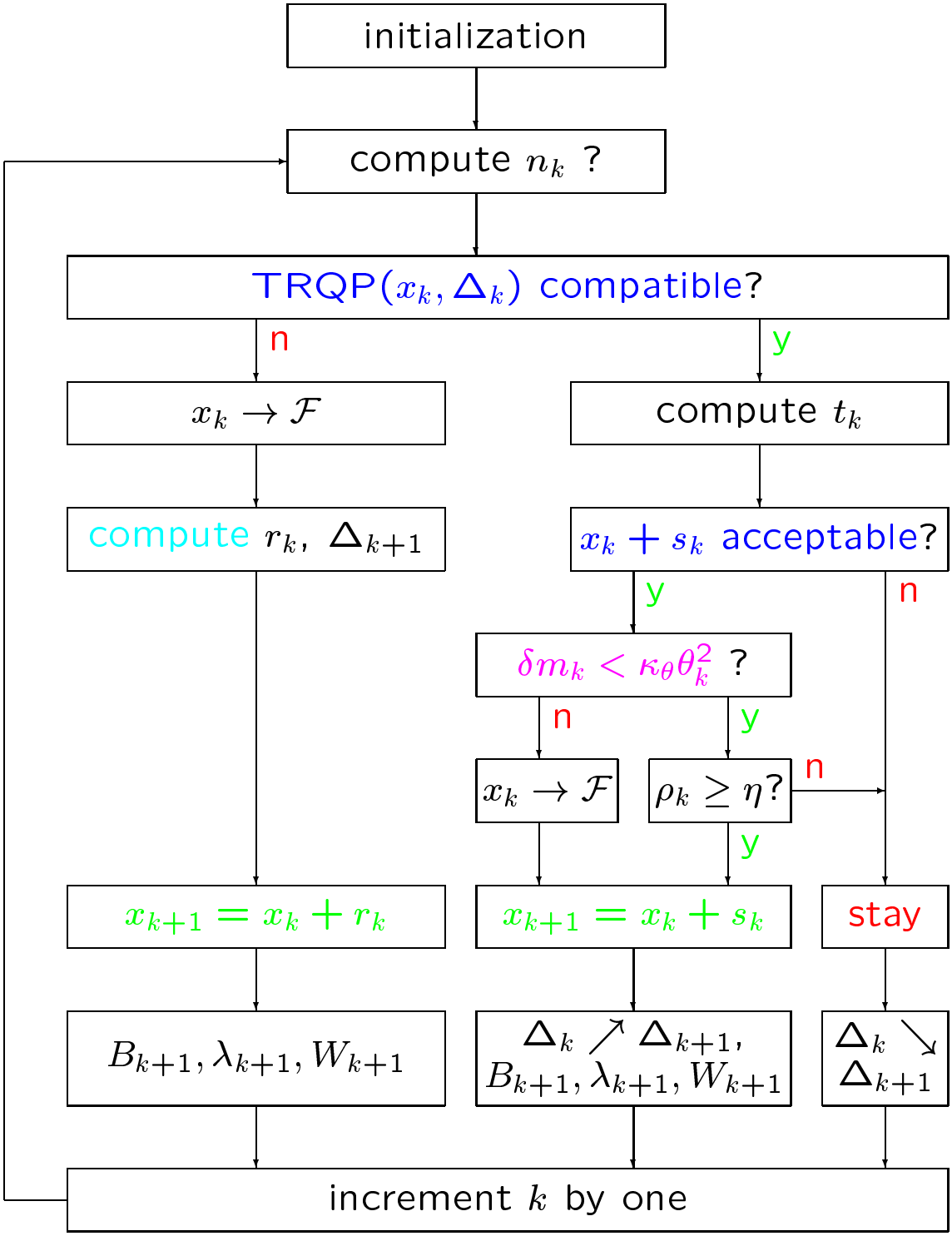
  $\Rightarrow$ step r_k (restoration step) ($k \in \mathcal{R}$)

Conditions:

- $x_k + r_k$ non-dominated, i.e. acceptable

- new TRQP($x_k + r_k, \Delta_{k+1}$) compatible

The SQP filter algorithm



The experience so far

- growing convergence theory
  (Fletcher, Leyffer, Gould, T, Ulbrich, Heinkenschloss, Vicente, Gonzaga, . . .)

- several algorithmic variants
  (merit function, non-monotone, IP)

- several experimental codes
  + FilterSQP (Fletcher, Leyffer)

- excellent numerical behaviour, especially on problems that are highly constrained

$\Rightarrow$ interest to extend the applicability of the idea

The feasibility problem

Feasibility

Find x such that

$$c(x) \geq 0$$

$$e(x) = 0$$

for general smooth c and e.

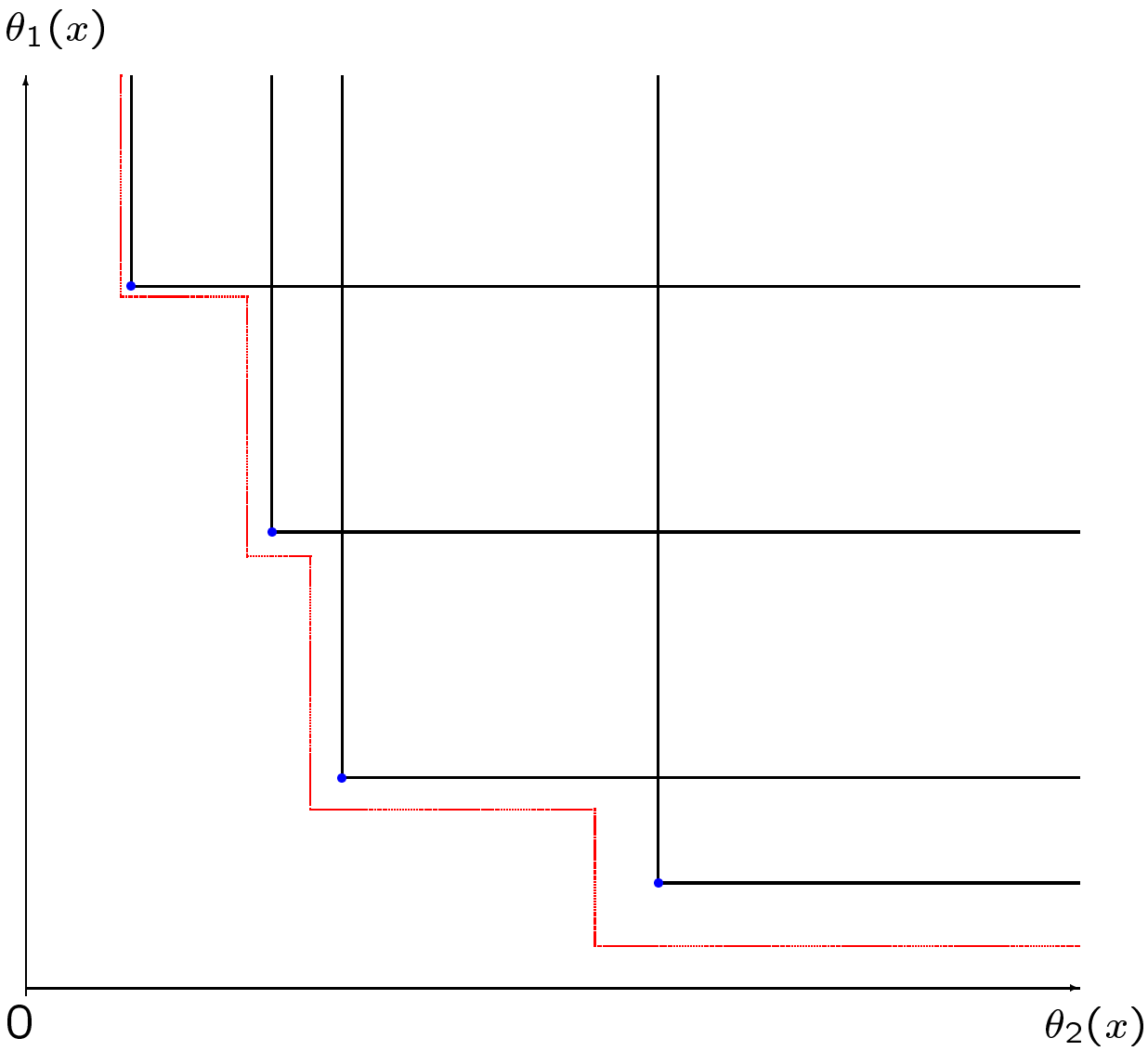
Least-squares

Find x such that

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

A mutidimensional filter algorithm

Idea: Use more dimensions for the filter space



(full dimension vs. grouping)

A mutidimensional filter algorithm (2)

Additionally

- compute the step s_k using
  the Gauss-Newton or the Newton model
  (or some adaptive combination)

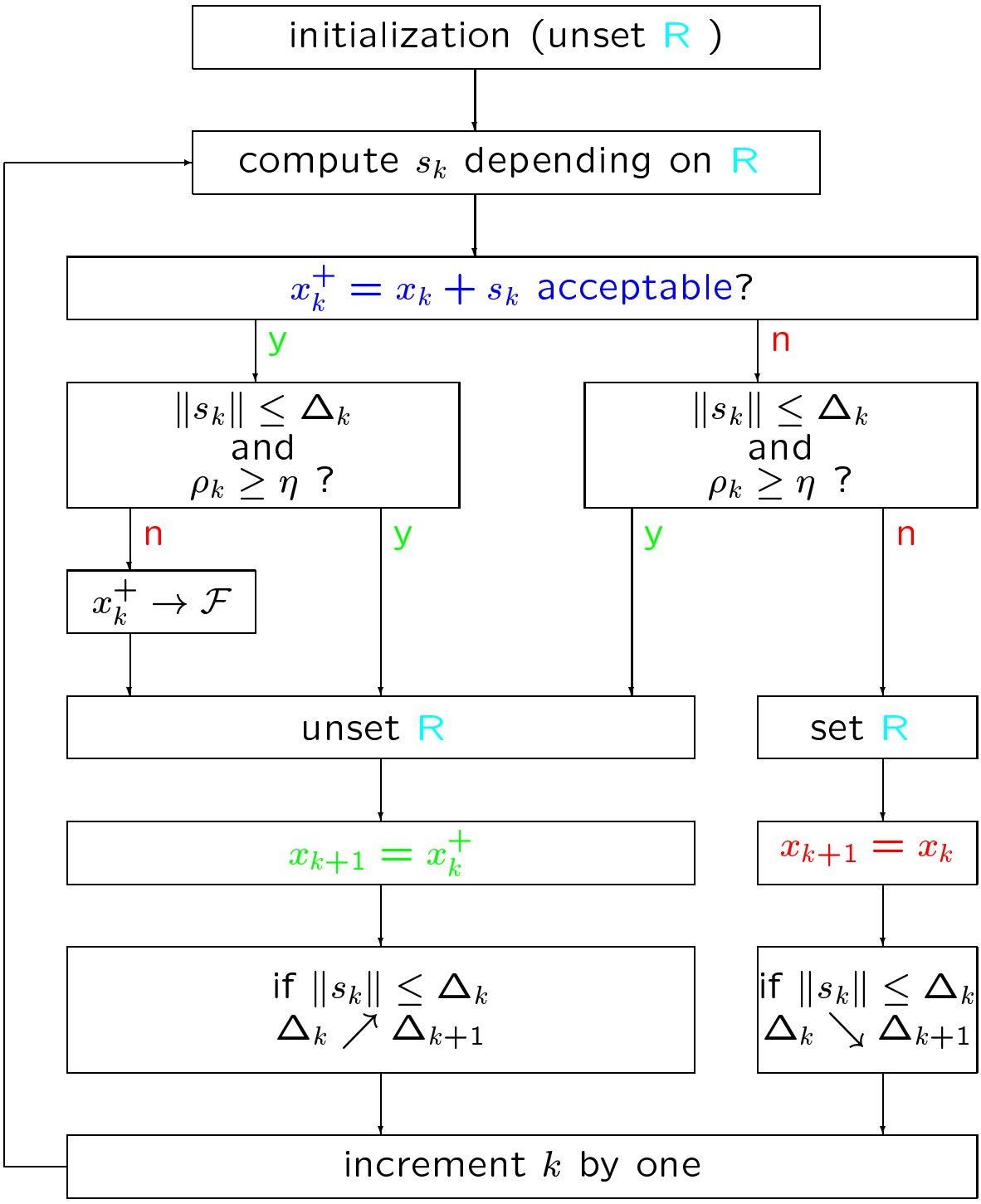
- possibly consider unsigned filter entries

- use TR algorithm when

  – trial point unacceptable

  – convergence to non-zero solution

The mutidimensional filter algorithm



A glimpse of the convergence theory

- the margin depends on $\|\theta\|$
- any subsequence of filter entries must converge to zero
- reduces to the ordinary trust-region convergence if only finitely many points in the filter

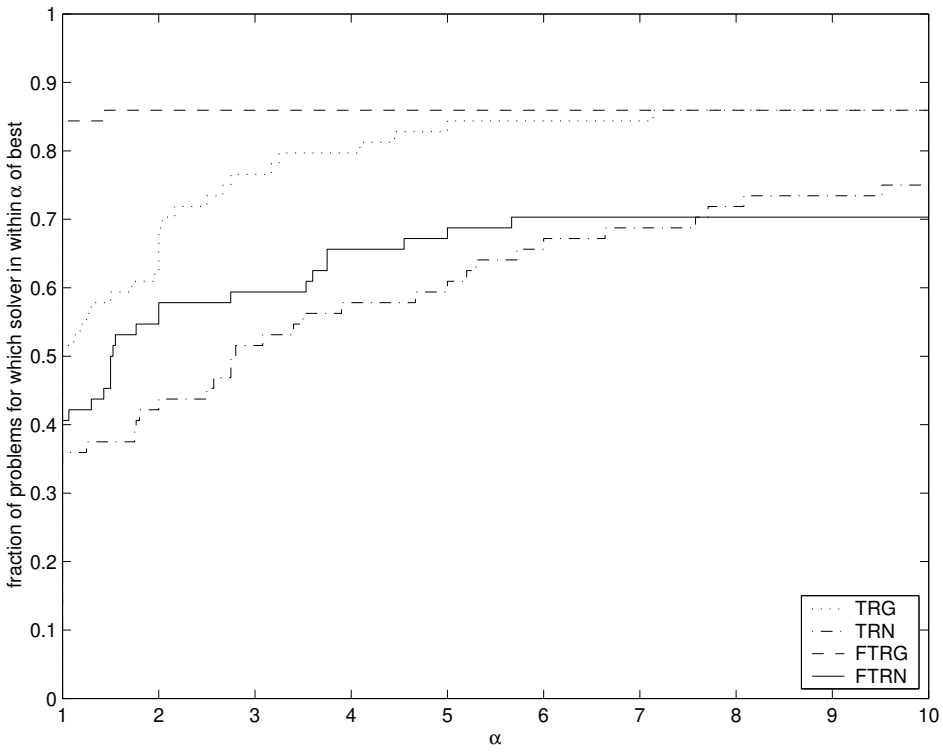
$\Rightarrow$ global convergence to

- either a zero of θ (feasible point)
- or a local (non-zero) minimum of $\|\theta\|$

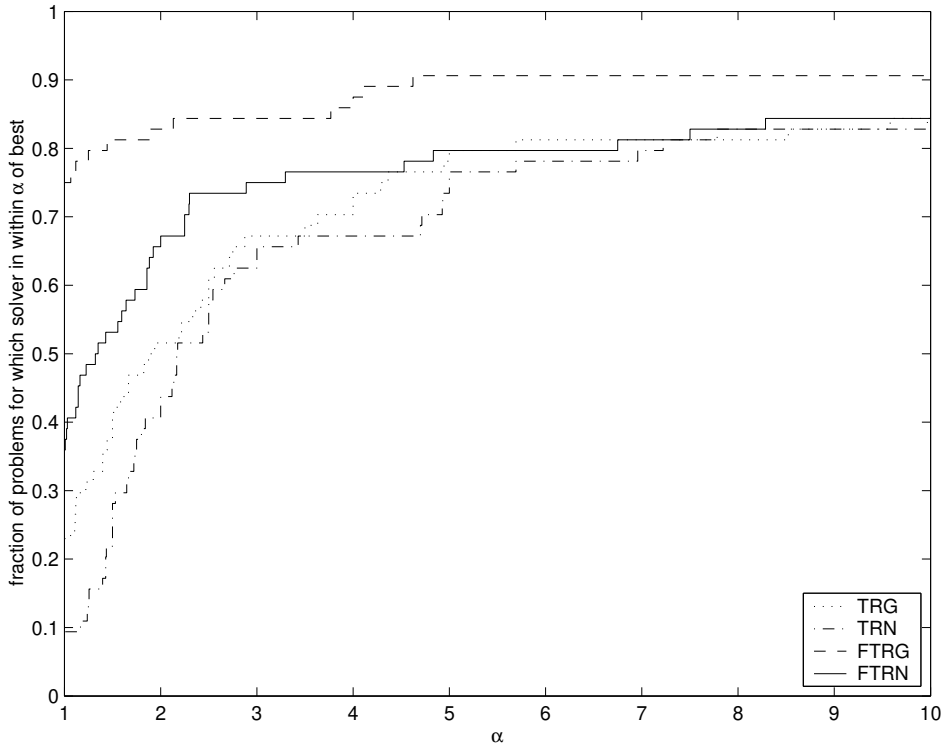
Numerical experience: FILTRANE

- 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
- will be part of the GALAHAD library

Numerical experience: trust-region vs. filter

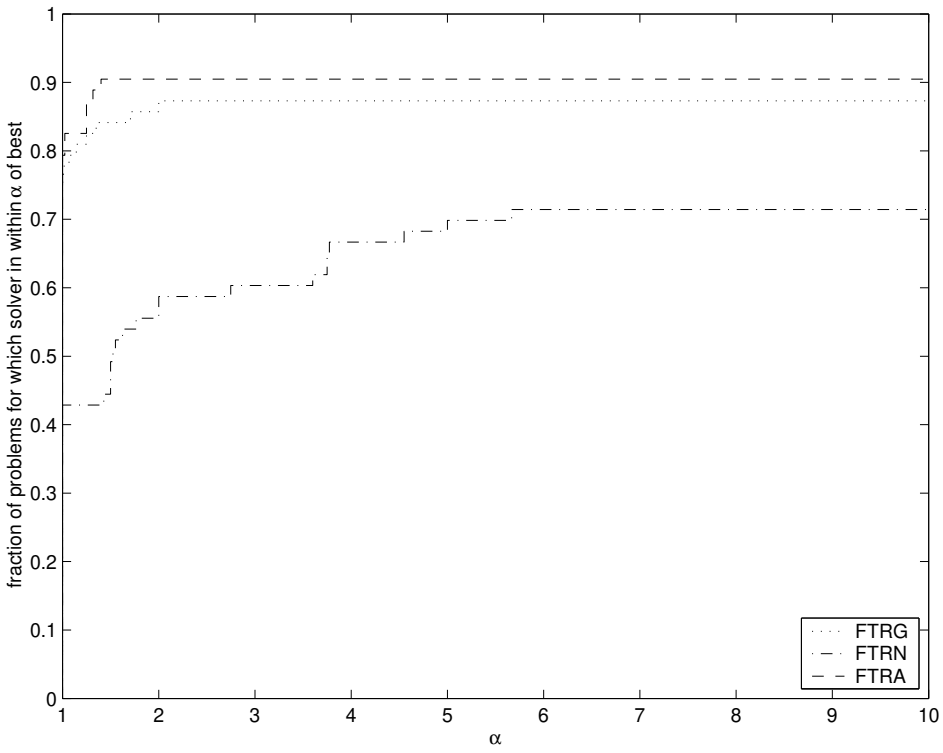


time

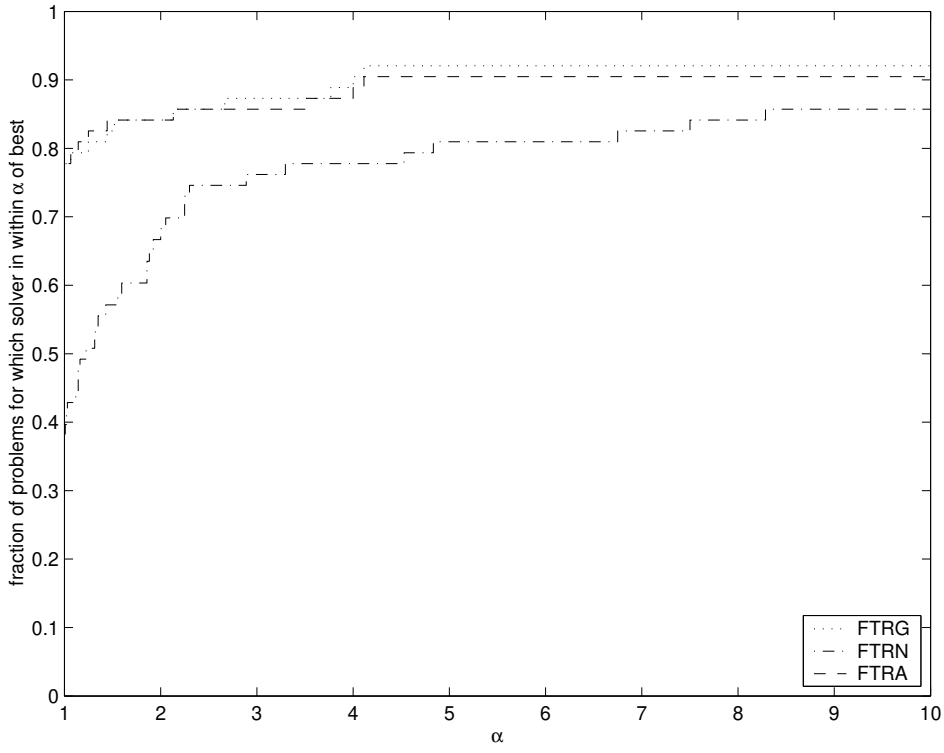


iters

Numerical experience: adaptive model

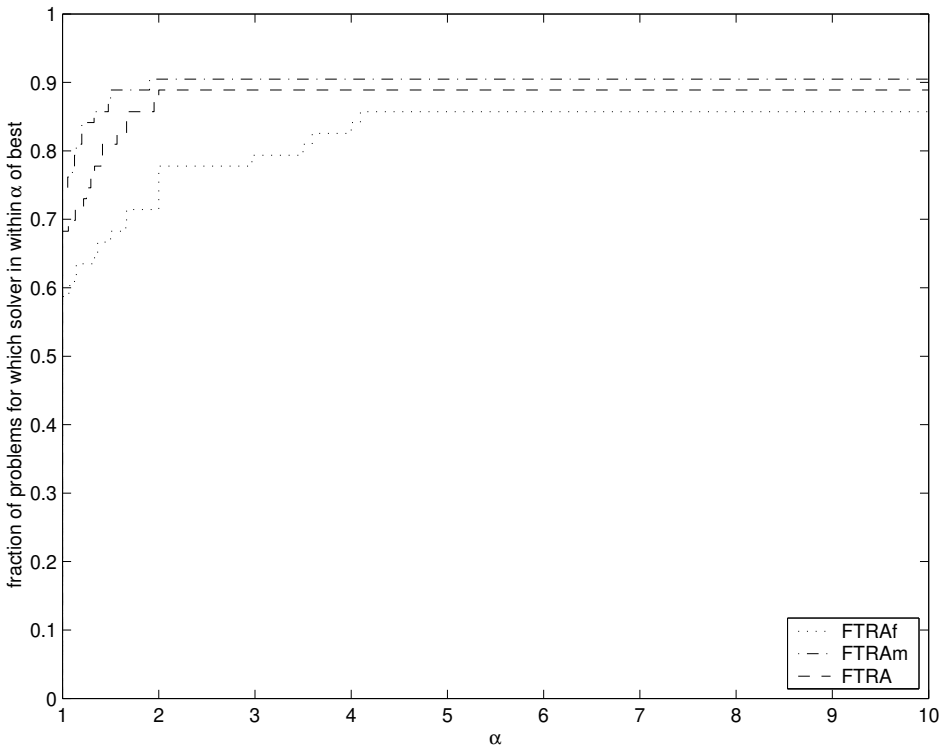


time

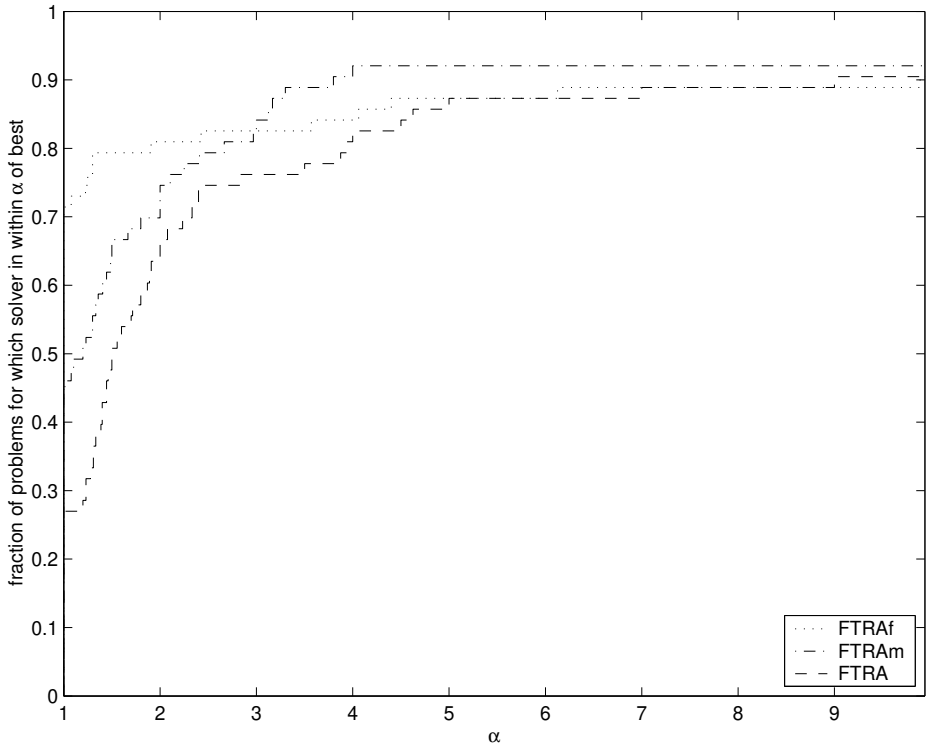


iters

Numerical experience: subproblem accuracy

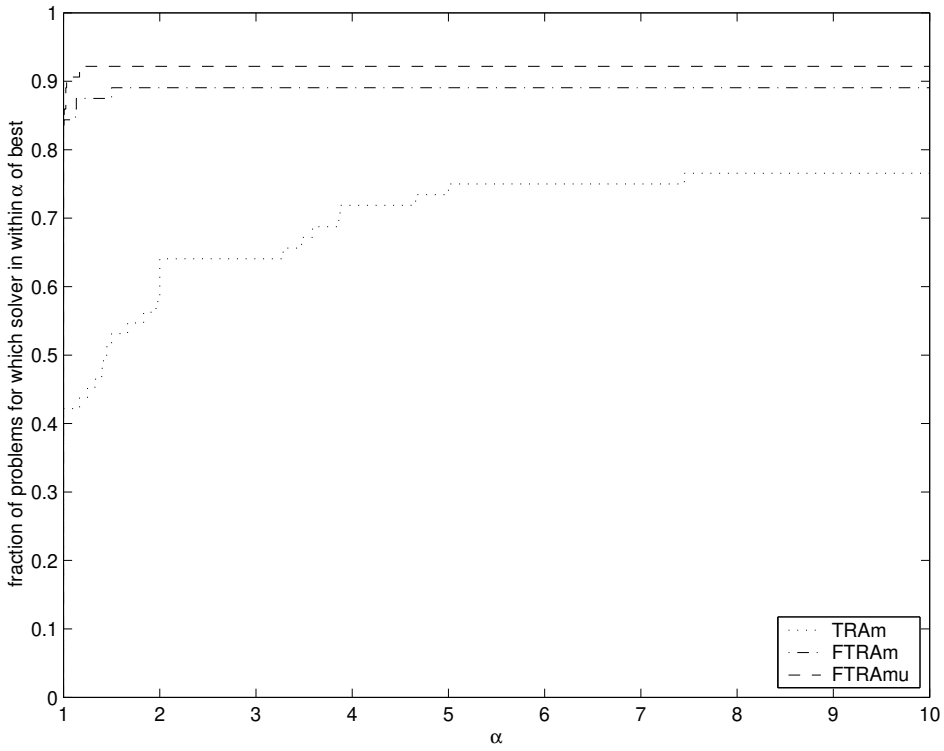


time

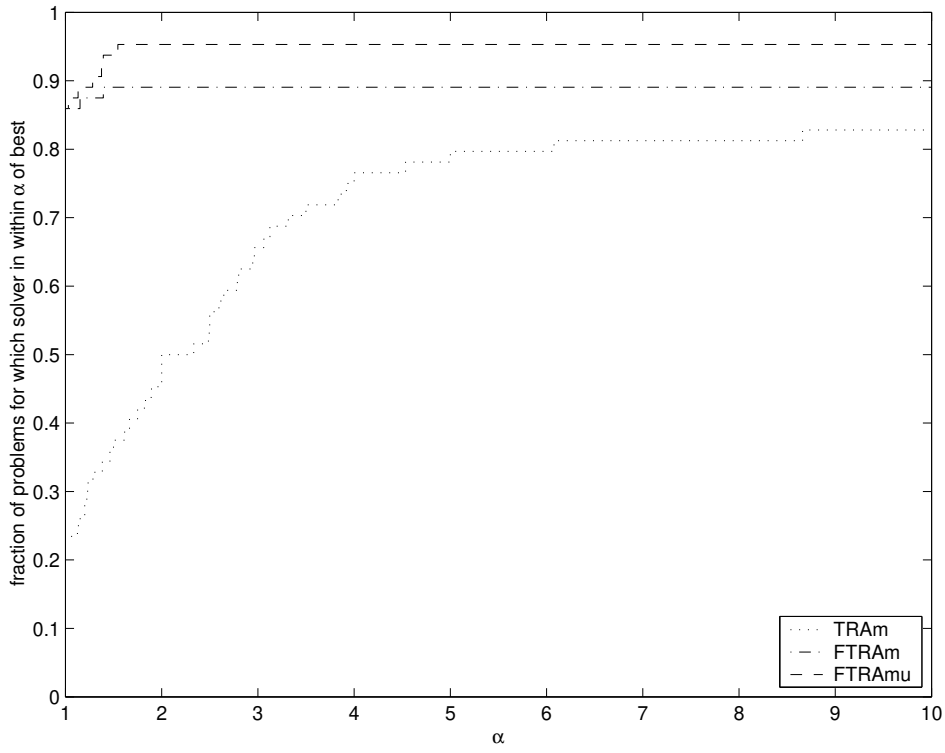


iters

Numerical experience: unsigned filter



time



iters

Conclusions

- encouraging numerical results
- sound global convergence theory
- illustrates power of non-monotone methods
- software soon available within GALAHAD
- use of multidimensional filter forthcoming in NLP and other areas

Thanks for your attention!