Sequential Quadratic Programming and Filter Methods

Ph. Toint
(with contributions by R. Fletcher, N. Gould and S. Leyffer)

FUNDP, Namur

The problem

The general nonlinear programming problem

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

for $x \in \mathbb{R}^n$, f and c smooth.
(today: emphasis on inequality constraints)

Many applications:

- *applied maths*: constrained fitting, optimal control, bi-level programming, statistics, ...
- *physics*: fluid flows, optics, geophysics,...
- *chemistry*: chem. engineering, kinetics, cristallography, molecular modelling,...
- *engineering*: structures, transportation, energy, circuit design, ...
- *computer science*: robotics, computer vision, neural networks, ...
- *biology and medecine*: NMR, pharmacokinetics, CA diagnostics, ...
- *economy*: utility models, portfolio management, quantitative game theory, ...

Conflicting objectives

At the same time:

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

$$\Rightarrow \text{CONFLICT}$$

Classical question: what is a better point?

Typically:

- compute a step s_k from some iterate x_k.
- is $x_k + s_k$ better than x_k?

A typical step from x_k

Minimize quadratic model $m_k(s)$ of the Lagrangian

$$\ell(x, y) = f(x) - y^T c(x)$$

subject to linearized constraints

(related to Newton step on
the 1-st order optimality conditions)

$\Rightarrow$ SQP step

QP(x_k):

$$\boxed{\begin{array}{ll} \text{minimize} & m_k(s) \\ \text{subject to} & c_k + A_k s \geq 0 \end{array}}$$

$\Rightarrow s_k$

Classical approach: merit functions

Need to balance feasibility and optimality

$\Rightarrow$ weighted sum

Define the maximum constraint violation :

$\theta(x) = \max[0, -c_1(x), \ldots, -c_m(x)] = \|c_-(x)\|_\infty$

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

(or similar: other norms, e^θ, . . .)

[external penalty function]

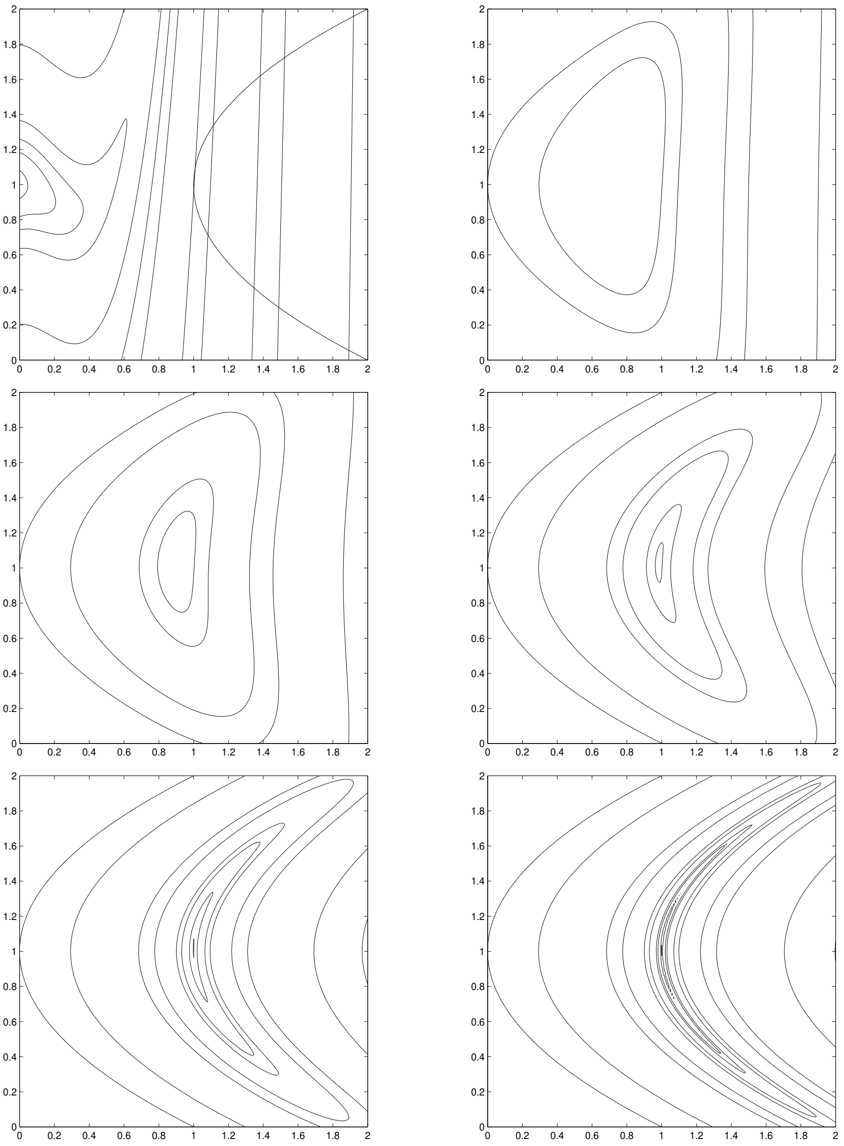
then (approximately) solve

$$\left\{\min_x \phi(x, \mu)\right\}_{\mu \to 0}$$

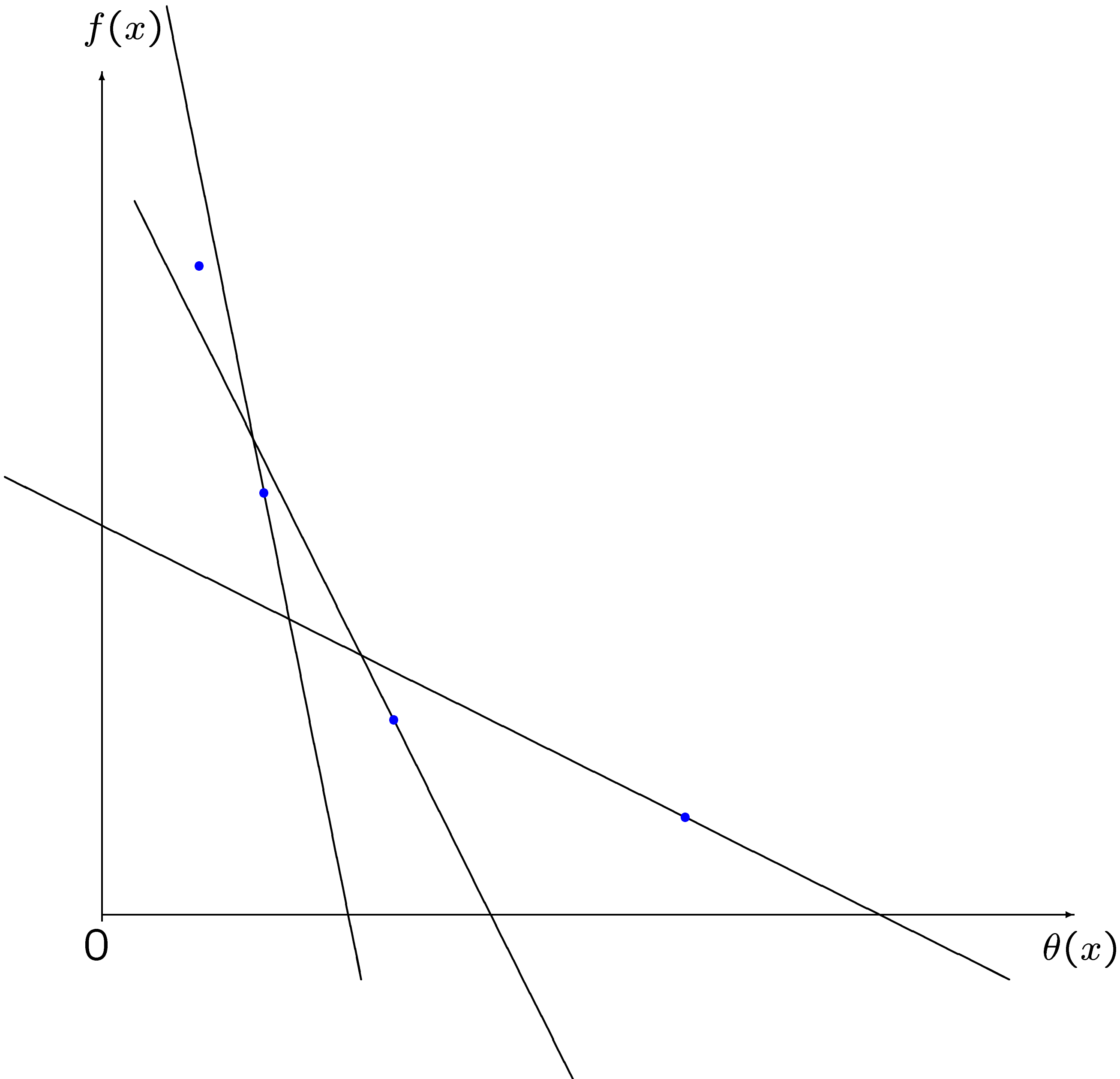
One proves (under reasonable conditions)

$$x(\mu) \to x_*$$

+ variations (augmented Lagrangians, . . .)


0 0.2 0.4 0.6 0.8 1 1.2 1.4 1.6 1.8 2
0 0.2 0.4 0.6 0.8 1 1.2 1.4 1.6 1.8 2

In the (f, θ) space



Classical approach: merit functions (2)

If x is feasible:

improve optimality while remaining feasible

$\Rightarrow$ barrier along the boundary

$$\boxed{\phi(x,\mu) = f(x) - \mu \log(\text{distance to infeas.})}$$

(or similar: $\mu(\text{distance})^{-1}$, ...)

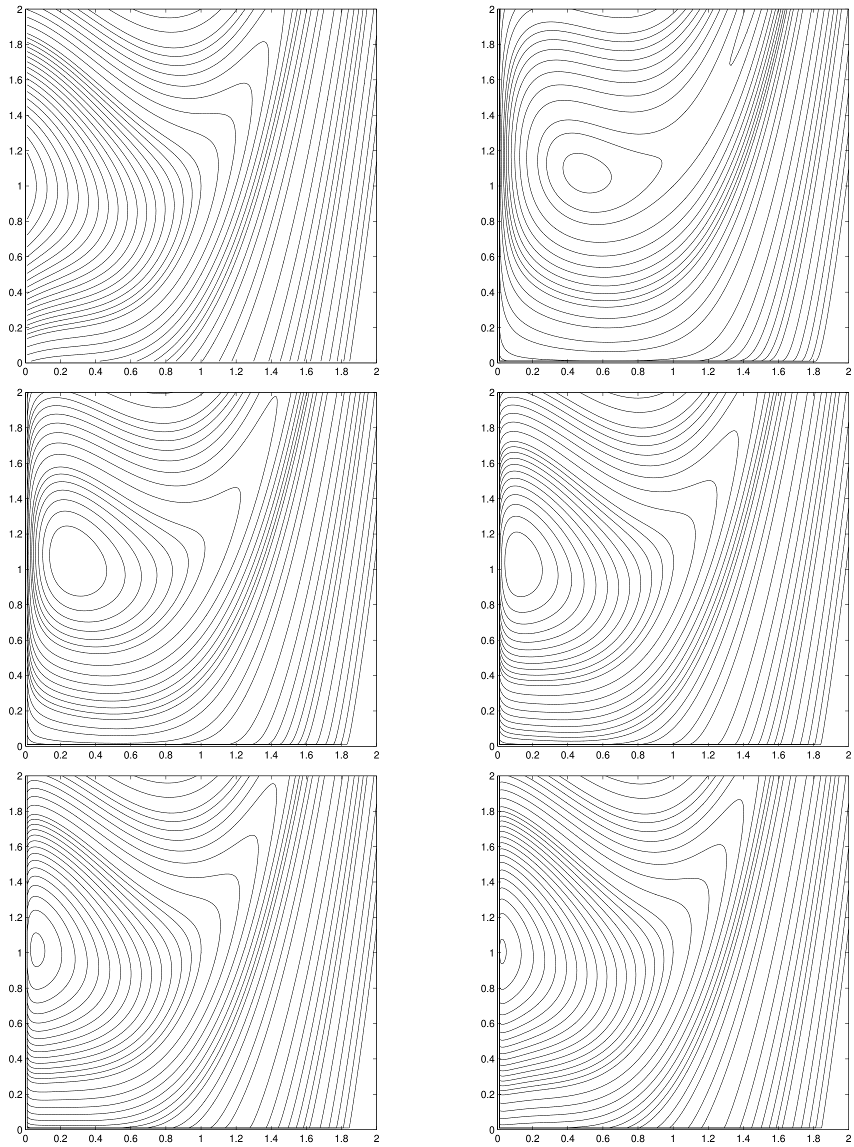
[internal penalty function = barrier function]

then (again) (approximately) solve

$$\left\{\min_x \phi(x,\mu)\right\}_{\mu\to 0}$$

One proves (under reasonable conditions)

$$x(\mu) \to x_*$$


0 0.2 0.4 0.6 0.8 1 1.2 1.4 1.6 1.8 2

Problems with merit functions

- the coherence between strategy to define a step (s_k) and the merit function $\phi(x, \mu)$ (Maratos effect)

- the practical rules to manage the penalty parameter μ
  - initial value ???
    (possible unbounded initial subproblem)
  - (non-)monotone updating rules . . .

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

A perspective shift

Replace classical question:

what is a better point?

by:

$\boxed{\text{What is a worse point?}}$

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)

$\gg$ concept from multi-objective optimization

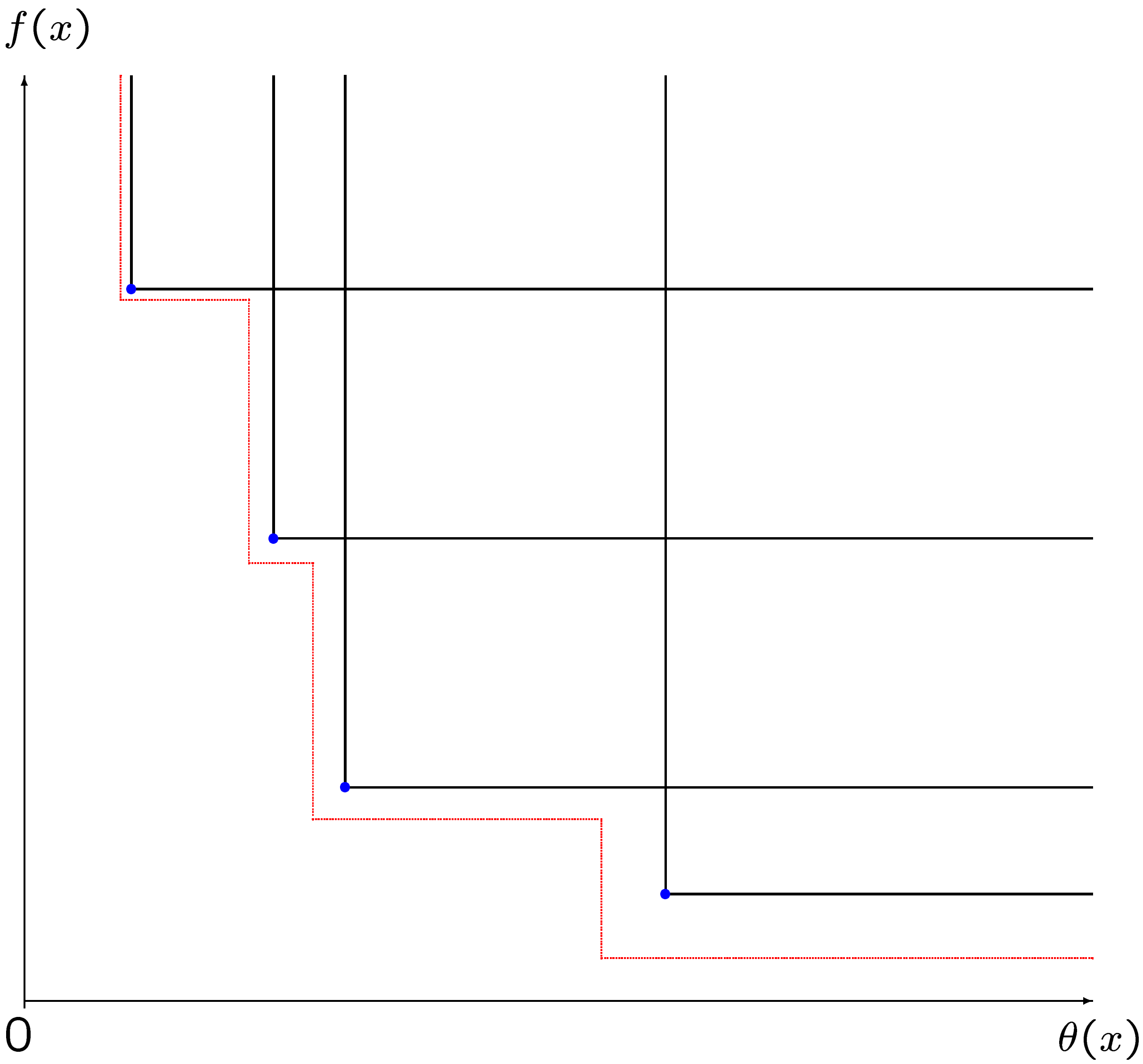
(Fletcher and Leyffer)

$\boxed{\text{When is } x_k + s_k \text{ acceptable?}}$

A Filter Algorithm

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

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



A Filter Algorithm (2)

Step selection:

- (approximate) SQP step
  - satisfies linearized constraints (when possible)
  - reduces a model m_k of the Lagrangian

  (more later on how to do this)

- impose trust-region to ensure global convergence (Δ_k)

TRQP(x_k, Δ_k):

$$
\begin{array}{ll}
\text{minimize} & m_k(s) = f_k + g_k^T s + \frac{1}{2} s^T H_k s \\
\text{subject to} & c_k + A_k s \geq 0 \\
\text{and} & \|s\|_? \leq \Delta_k,
\end{array}
$$

The basic idea of trust regions

A very general idea:

$m_k(s)$ based on Taylor expansion

$\Rightarrow$ only accurate locally !!!

$\Rightarrow$ limit investigation to suitable
neighbourhood of x_k (trust region)

$\Rightarrow$ iteratively adapt size of the region
(based on quality of prediction)

Supported by very strong global convergence theory (outside filter algorithms)

. . . but linesearch filter also possible?

A Filter Algorithm (3)

Two types of steps:

- steps that reduce infeasibility:
  keep track of progress by inserting points in the filter

- steps that reduce the objective function
  mimic an unconstrained (trust-region based) minimization

Criterion :

$$\boxed{\delta m_k \geq \kappa \theta_k^2}$$

A Filter Algorithm (4)

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) ($k \in \mathcal{R}$)

- insert failed point in the filter

At exit :

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

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

A Filter Algorithm (5)

Why are we interested?

- excellent performance in practice
  competitive with best codes (Nocedal)

- code (filterSQP) available from Dundee

- experimentation possible on NEOS

$\Rightarrow$ crucial to

- optimize the algorithm(s)

- understand why it works so well

Which SQP step?

Strategy 1 global solution of TRQP(x_k, Δ_k)

choose s_k at the global solution of

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

- conceptually simple
- covered by theory: global convergence to first-order critical points (for SLP: Fletcher, Leyffer and T., 1998, for SQP: Fletcher, Leyffer and T., 2000)
- adequate for convex problems
- irrealistic for nonconvex ones
- potentially costly

Which SQP step?

Strategy 2   a composite step

1. choose a normal step n_k that satisfies linearized constraints, with
   θ_k small enough $\Rightarrow \|n_k\| = O(\theta_k)$

   (when possible, otherwise $\rightarrow$ restoration),

2. choose a tangential step t_k from
   $x_k^N = x_k + n_k$ that reduce m_k

$$\Rightarrow s_k = n_k + t_k$$

- realistic for convex and nonconvex problems

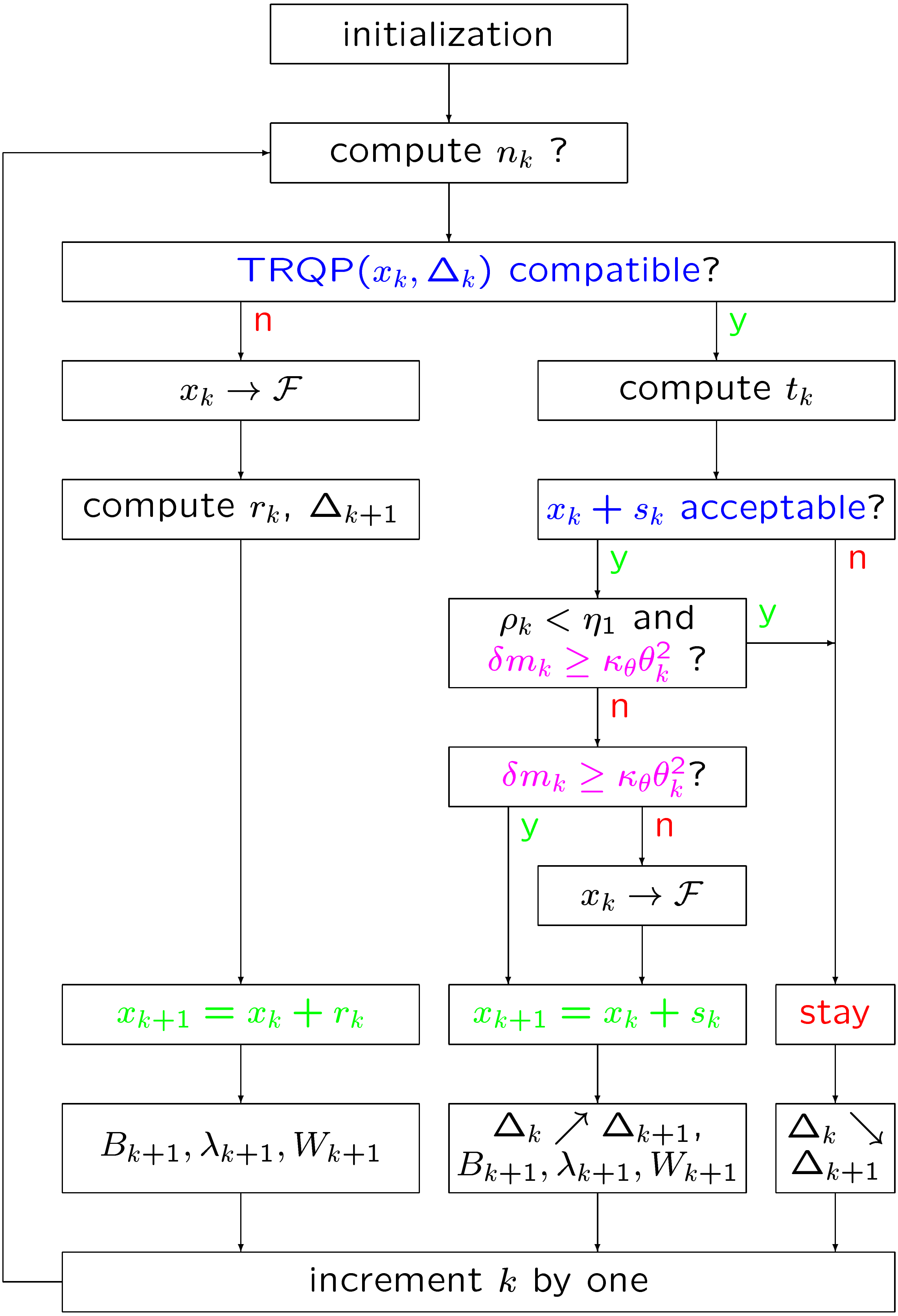
- allows for approximate solution of TRQP: only a Cauchy minimum decrease is requested in m_k

$$\delta m_k^t \geq \kappa \pi_k \min\left[\frac{\pi_k}{\kappa}, \Delta_k\right]$$

- covered by theory: global convergence to first-order critical points (Fletcher, Gould, Leyffer, T., 2000)

- moderately complex

- potentially costly (...but approximation may help)


initialization
compute n_k ?
TRQP(x_k, Δ_k) compatible?
n
y
$x_k \to \mathcal{F}$
compute t_k
compute r_k, Δ_{k+1}
$x_k + s_k$ acceptable?
y
n
$\rho_k < \eta_1$ and $\delta m_k \geq \kappa_\theta \theta_k^2$?
y
n
$\delta m_k \geq \kappa_\theta \theta_k^2$?
y
n
$x_k \to \mathcal{F}$
$x_{k+1} = x_k + r_k$
$x_{k+1} = x_k + s_k$
stay
$B_{k+1}, \lambda_{k+1}, W_{k+1}$
$\Delta_k \nearrow \Delta_{k+1}$, $B_{k+1}, \lambda_{k+1}, W_{k+1}$
$\Delta_k \searrow \Delta_{k+1}$
increment k by one

Outline of convergence theory (1)

Assumptions:

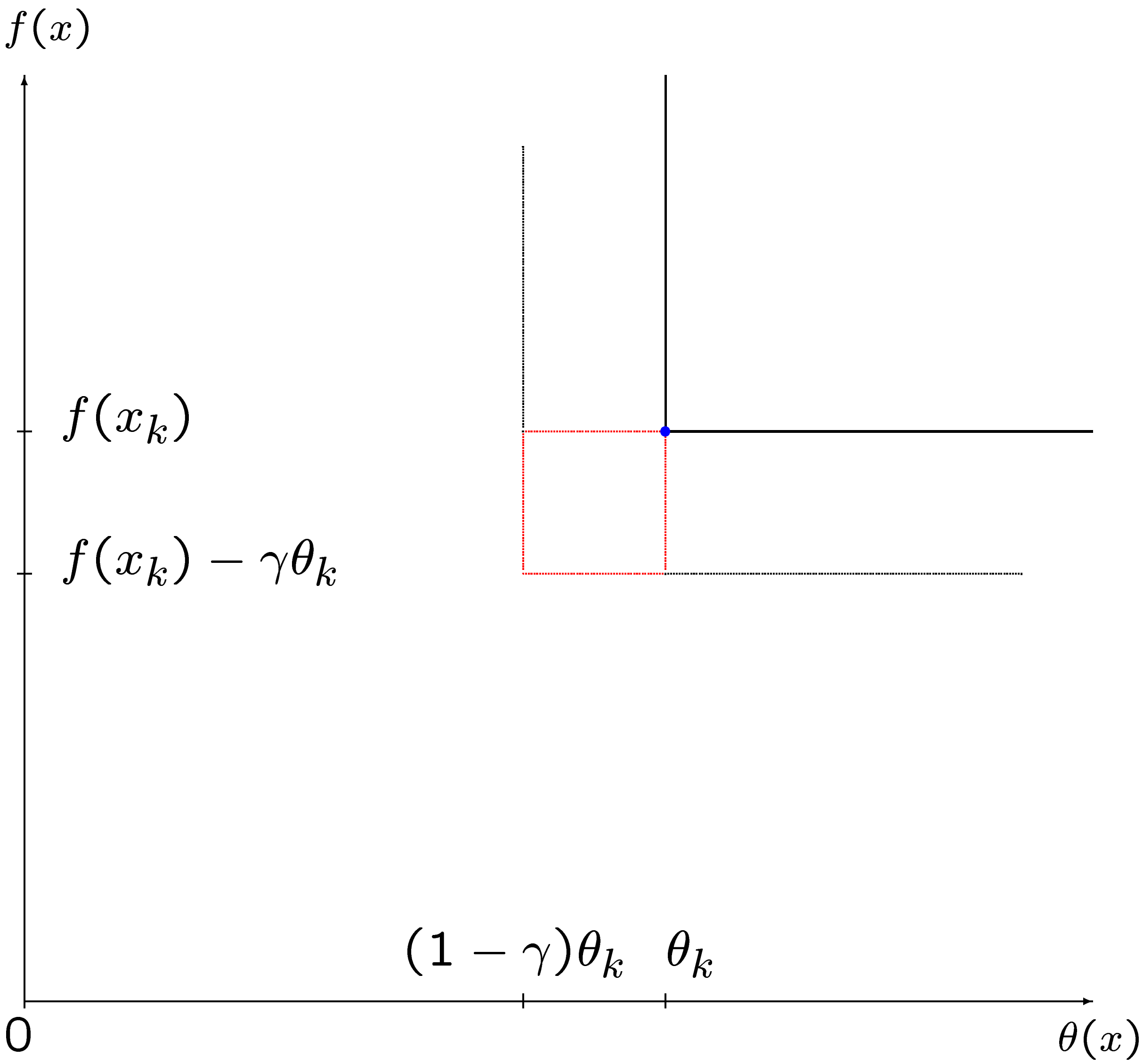
- $f(x)$ and $c(x)$ smooth
- $\{x_k\}$ bounded
- $\{H_k\}$ bounded

$$\boxed{\begin{array}{c} \{k_i\} \text{ infinite and } x_{k_i} \to \mathcal{F} \\ \Rightarrow \lim_{i\to\infty} \theta_{k_i} = 0 \quad (\text{feasible}) \end{array}}$$

(filter fills up)

Filling up the filter

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



$\Rightarrow$ finitely many such points possible
(note: inclusion property)

Outline of convergence theory (2)

$k \notin \mathcal{R}$, θ_k and Δ_k small enough and x_k not critical

$$\Rightarrow \delta m_k \geq \kappa\epsilon\Delta_k \quad \text{(linear decrease)}$$

$$\Rightarrow \rho_k \geq \eta_2 \quad \text{(success)}$$

$$\Rightarrow \delta m_k \geq \kappa_\theta \theta_k^2 \quad \text{(tangential)}$$

$$\Rightarrow f(x_k + s_k) \leq f(x_k) - \gamma_\theta \theta_k$$

θ_k and Δ_k small enough and x_k not critical

$$\Rightarrow k \notin \mathcal{R}$$

$\{k_i\}$ infinite and $x_{k_i} \to \mathcal{F}$

$$\Rightarrow \{x_{k_i}\} \to \text{critical point}$$

Outline of convergence theory (3)

Remains: $\{k_i \mid x_{k_i} \to \mathcal{F}\}$ finite

$\Rightarrow \lim_k \theta_k = 0$ (feasible)

$\Rightarrow$ classical convergence theory

- x_k not critical $\Rightarrow \Delta_k$ bounded below
- $\{x_{k_i}\} \to$ critical point

Which SQP step?

Strategy 3   a LP-based hybrid step

Idea : first perform a step which ensures global convergence, then improve upon it with a locally good step

$\rightarrow$ Cauchy step philosophy

- globally convergent step:
  SLP step (solve TRQP with $H_k = 0$)

- attempt to include second order behaviour
  - solve Equality constrained QP (EQP) (drop constraints that are inactive at the SLP solution)
  - second-order correction

  $+$ test if improvement obtained

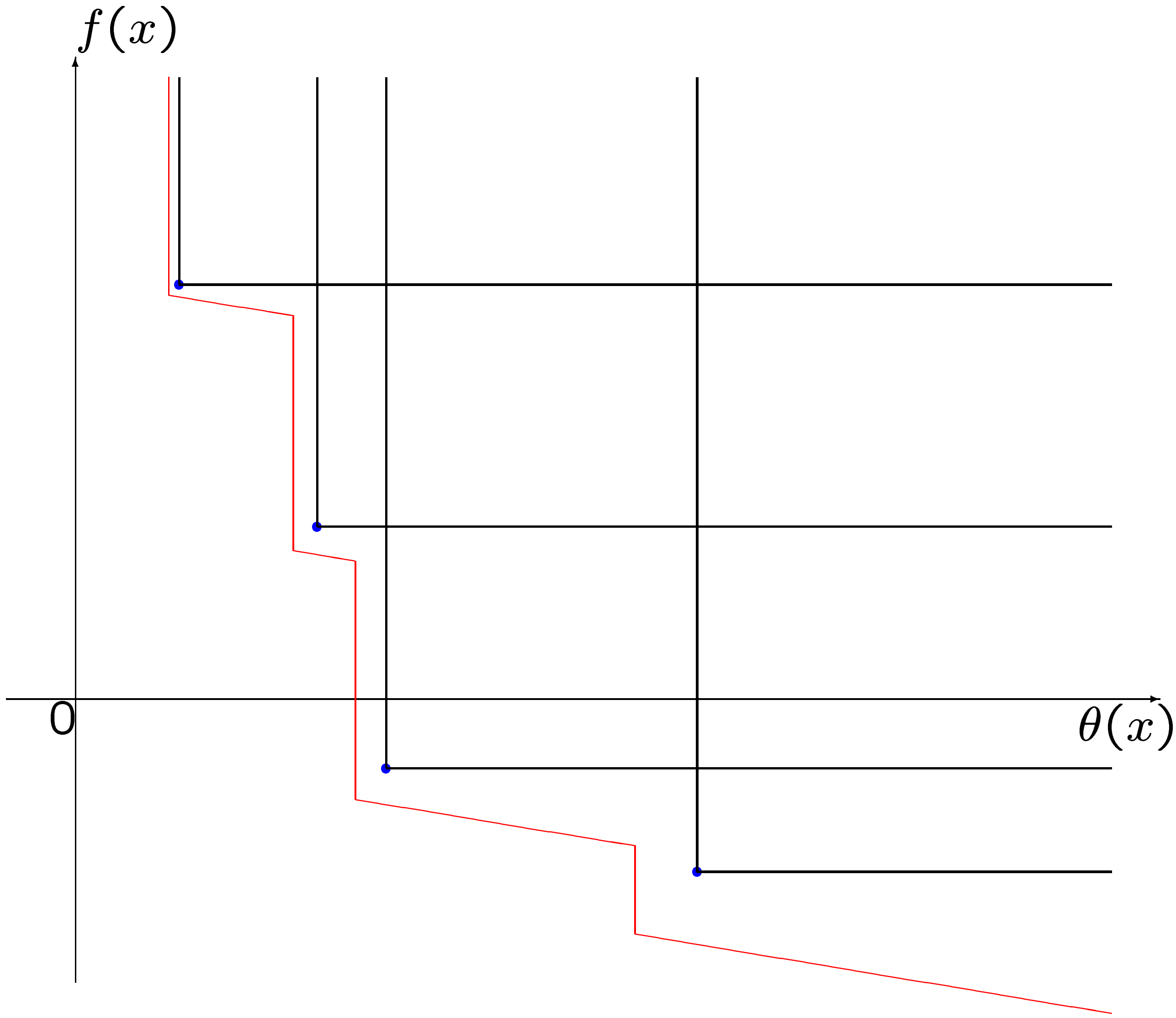
- “natural” idea

- covered by theory: global convergence to first-order critical points (Fletcher and Chin, 2001)
  $\rightarrow$ use of a slanting enveloppe

- may allow second-order correction steps

- potential for linear behaviour

The slanting filter enveloppe



(variant on the inclusion property)

Which SQP step?

$\boxed{\text{Strategy 4}}$ a QP-based hybrid step

Idea : first attempt a good local step, test it for global convergence properties and fall back on the composite step if necessary

- test uses a global criticality measure π_k

- explicitly test a Cauchy-like condition

$$\delta m_k \geq \kappa \min\left[\pi_k, \Delta_k\right]$$

- compute the composite step if fails

- exploits Newton step as much as possible

- covered by theory: global convergence
  to first-order critical points
  (Gould and Toint, 2001)

- potentially costly
  (if first steps fails often)

Not discussed in this talk

- Filter and interior-point methods:
  optimality $\leftrightarrow$ centering
  (Heinkenschloss and Vicente, 2000)

- Use of filter ideas in derivative free (DFO) optimization algorithms:

  - SQP-like + multivariate interpolation:
    Colson and T., 2001

  - constrained direct search techniques:
    Audet and Dennis, 2000.

- links with other non-monotone constrained optimization
  (e.g. Ulbrich and Ulbrich, 2000)

- . . . (?)

Conclusion

- a very active field of research
- adequate first-order convergence theory for different SQP variants
- diversity indicates that the field is not mature
  $\Rightarrow$ a lot more work to do. . .

Ongoing:

- filter for nonlinear equations
- local convergence issues
  - active constraint identification
  - rate of convergence