Recursive Trust-Region Method for Unconstrained Minimization: Convergence properties and experience

Philippe Toint (philippe.toint@fundp.ac.be)

(with S. Gratton and A. Sartenaer)

Department of Mathematics, University of Namur, Namur

Unconstrained optimization

The unconstrained nonlinear programming problem:

$$\text{minimize} f(x)$$

for $x \in \mathbb{R}^n$, f smooth.
Many applications:

- variational problems
- surface design

(Generalizations now under study)

Hierarchy of problem descriptions

Can we use a structure of the form:

	Finest problem description	
Restriction ↓ R		P ↑ Prolongation
	Fine problem description	
Restriction ↓ R		P ↑ Prolongation
	...	
Restriction ↓ R		P ↑ Prolongation
	Coarse problem description	
Restriction ↓ R		P ↑ Prolongation
	Coarsest problem description	

Structured model choice

Consider minimizing at topmost (finest) level.
At each iteration, choose the model as

- a local Taylor expansion (classical)
  $\longrightarrow$ Taylor iteration
- the immediately coarser problem description
  $\longrightarrow$ recursive iteration:

compute fine g (and H)		step and trial point
Restriction $\downarrow$ R		P $\uparrow$ Prolongation
minimize the *coarse* model within the *fine* TR		

A recursive multi-scale algorithm

Until convergence:

- Choose either a Taylor or recursive model
  - Taylor model: compute a Taylor step
  - Recursive: apply the Algo recursively
- Evaluate change in the objective function
- If achieved reduction $\approx$ predicted reduction,
  - accept trial point as new iterate
  - (possibly) enlarge the trust region

  else
  - reject the trial point
  - shrink the trust region
- Impose: current TR $\subseteq$ upper level TR

The University of Namur
www.fundp.ac.be

(Simple) Taylor iterations

Which solver for the (approximate) solution of the (same level) trust-region subproblem?

Simple answer:

- for low(est) level(s) (small dimension):
  the exact Moré-Sorensen method
- for higher levels (high dimension):
  a truncated conjugate gradient
  (Steihaug-Toint or GLTR)

But. . .

Multigrid technology...

Multigrid techniques for PDE problems indicates:

- the high-frequency components of residual only visible in fine mesh (high levels)
- need two different methods:
  - reduce high frequency components on the fine mesh

    Smoothing

  - reduce low frequency components on the coarse mesh

    Damping

…adapted to optimization

In unconstrained optimization,

residual → gradient

- gradient smoothing:
  - TCG not very efficient!
  - adapt Gauss-Seidel smoothing
    → cyclic coordinate search
    (on Taylor's model)
- low frequency damping:
  full solution (MS) in low dimension

Structure of the recursive iterations

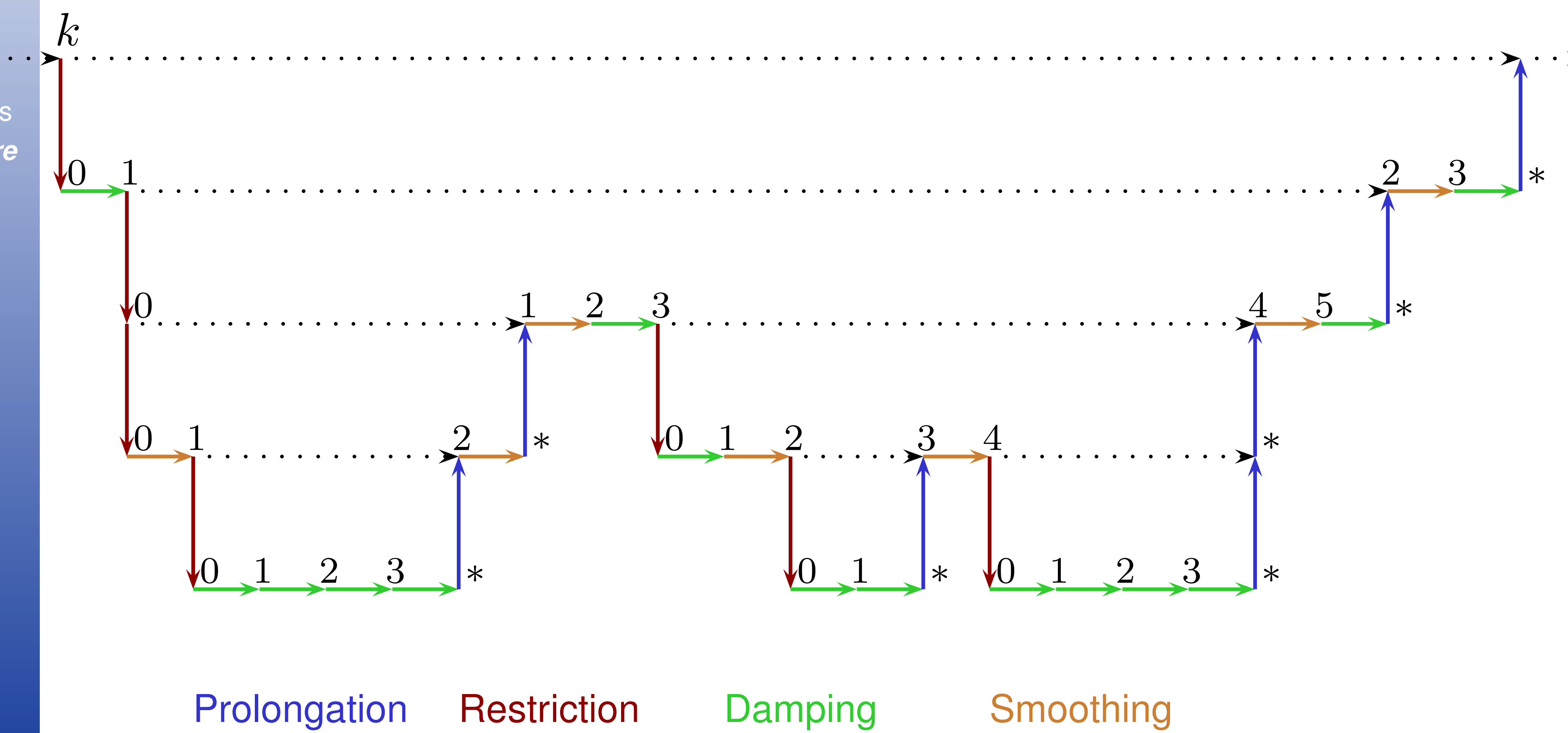
Decision to stop solving the lower-level subproblem based on

- subproblem criticality $\longrightarrow$ free form (gradient accuracy + TR constraint activity)
- fixed form cycles (possibly truncated)
  - V cycles
  - W cycles
  - Wq cycles ($q > 2$)

At least one successful iteration per level

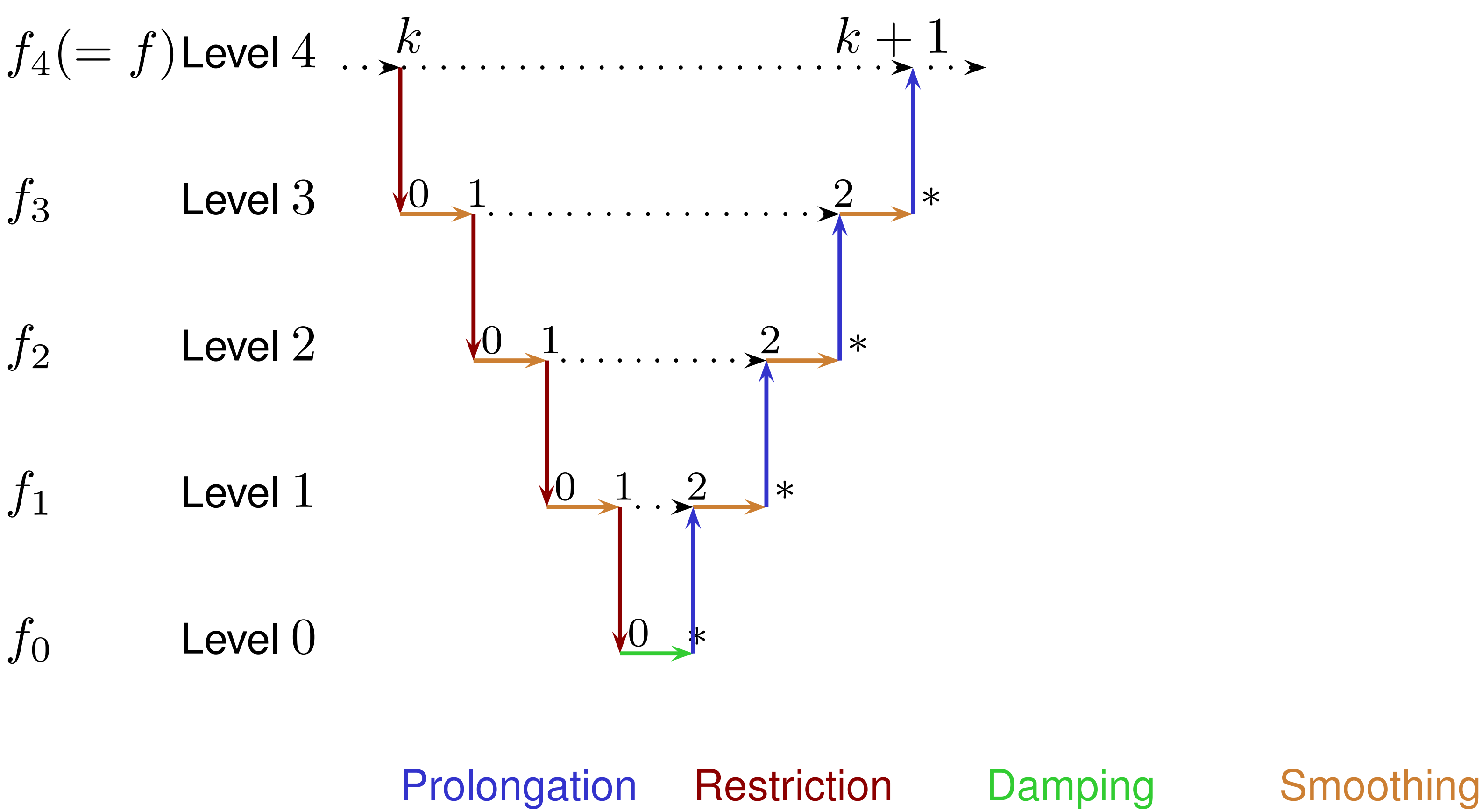
Free form iterations

The “iterates view” for the same example:



V cycles

The "iterates view" for a V cycle recursion (5 levels, all iterations successful):



Computing the initial point

Need $x_{r,0}$ (starting point at topmost level):
$\rightarrow$ use a mesh refinement technique.

For $i = 0, \ldots, r-1$,

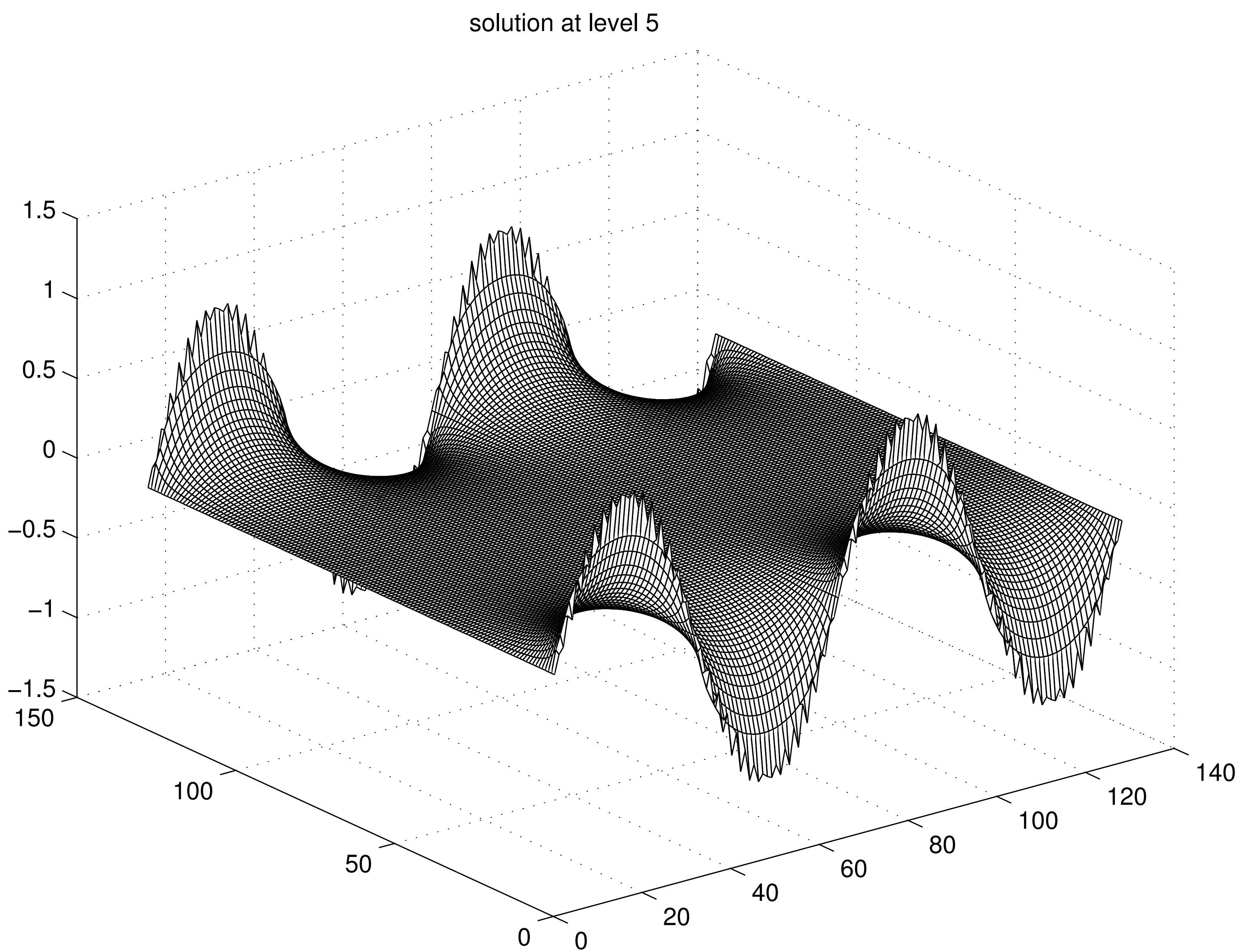
- apply the recursive algorithm to solve

$$\min_x f_i(x)$$

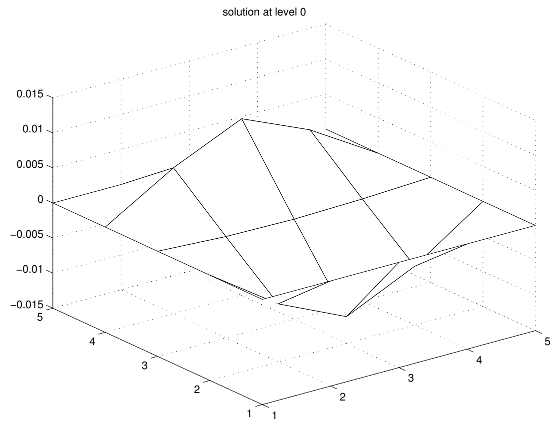
  (with increasing accuracy)
- apply the prolongation to obtain the initial point at next level

- reminiscent of the full multigrid scheme
- approach of the solution at coarse levels

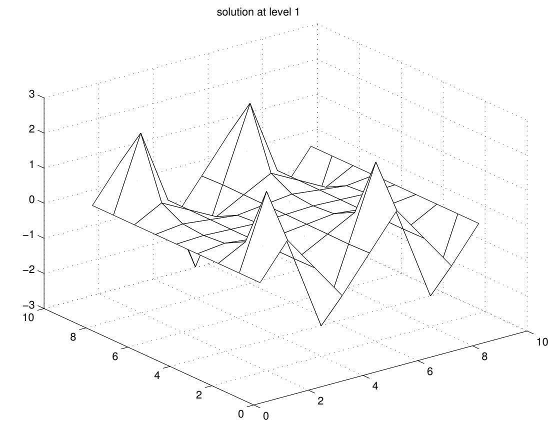
A minimum surface problem



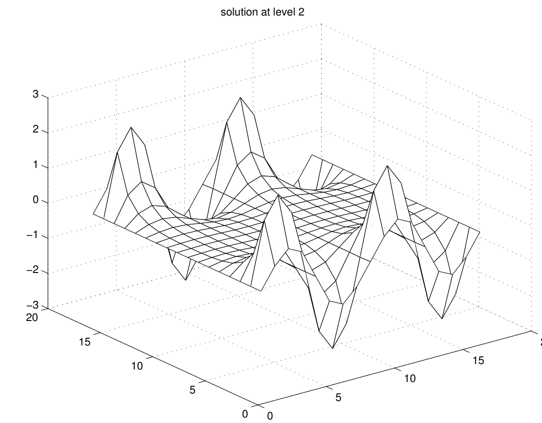
The level structure



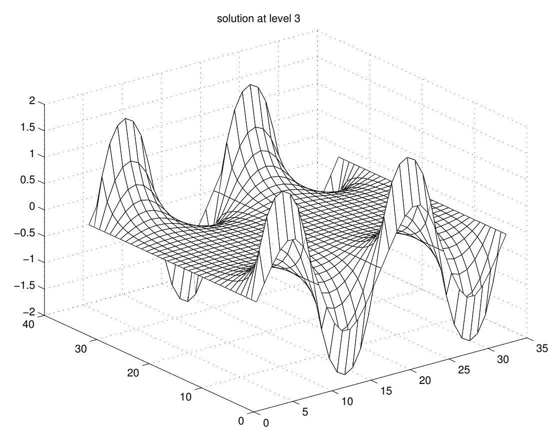
$n = 3^2 = 9$



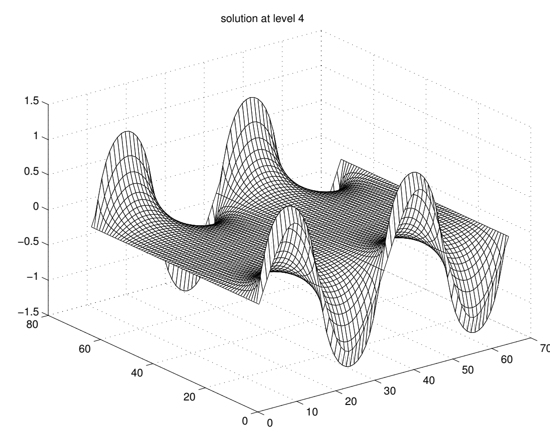
$n = 7^2 = 49$



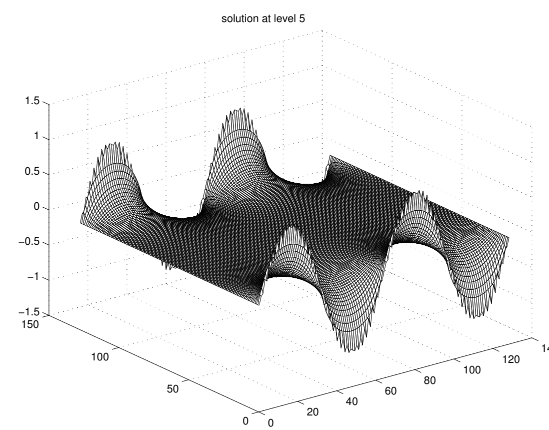
$n = 15^2 = 225$



$n = 31^2 = 961$



$n = 63^2 = 3969$



$n = 127^2 = 16129$

Further problem details

- structured level transfer operators
  - $P =$ full weighting interpolation operator
  - $R =$ normalized P^T
- handling the boundary condition
  - boundary condition not forced
  - additional smoothing "just inside"
- random starting point (at coarsest level)

Contact me for a live demo . . .

Other test problems (1)

- 2D Laplacian (check) problem (5 points FD pencil, unit square)

$$\min -\tfrac{1}{2}x^T \Delta x - f^T x$$

$$f = \sin[x_1 * (1 - x_1)] * \sin[x_2 * (1 - x_2)]$$

- 2D nonconvex quartic nonlinear least-squares (5 points FD pencil, unit square)

$$\min \int (u-f)^2 + 10^{-2} \int (\gamma - f)^2 + \int (-\Delta u + u\gamma - g)^2$$

$$g = -\Delta f + f^2$$

Other test problems (2)

- Lewis and Nash Dirichlet to Neumann transfer problem:

$$\min_{a\,:\,[0,\pi]\to\mathbb{R}} \int_0^\pi \left(\frac{\partial u}{\partial x_2}(x_1,0) - \phi(x_1) \right)^2 dx_1$$

with
$$S = \{(x_1,x_2), 0 \le x_1 \le \pi, 0 \le x_2 \le \pi\}$$
$$\Gamma = \{(x_1,x_2), 0 \le x_1 \le \pi, x_2 = 0\}.$$
$$\phi(x) = \sum_{i=1}^{15} \sin(i\,x) + \sin(40\,x)$$

subject to the boundary value problem

$$\begin{cases} \Delta u = 0 \\ u(x,y) = a(x_1) \text{ on } \Gamma, \quad u(x,y) = 0 \text{ on } \partial S \backslash \Gamma \end{cases}$$

Other test problems (3)

- Borzi and Kunish's solid ignition optimal control (unit square):

$$\min_f \left\{ \int_\Omega (u-z)^2 + \frac{\beta}{2}\int_\Omega (e^u - e^z)^2 + \frac{\nu}{2}\int_\Omega f^2 \right\}$$

where

$$\begin{cases} -\Delta u + \delta \exp(u) = f \text{ in } \Omega \\ u = 0 \text{ on } \partial\Omega. \end{cases}$$

Other test problems (4)

- Vogel's image deblurring inverse problem:

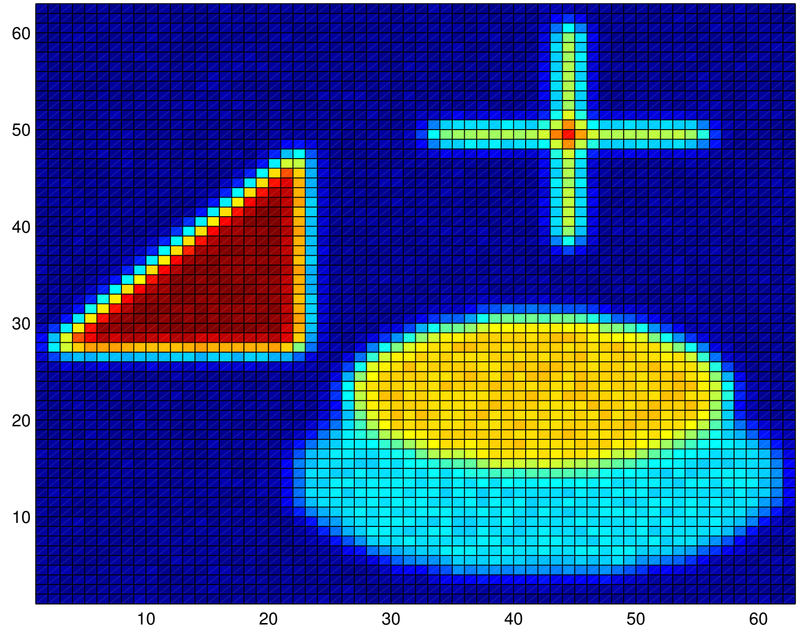
$$\min \mathcal{J}(f) = \frac{1}{2}\|Tf - d\|_2^2 + TV(f),$$

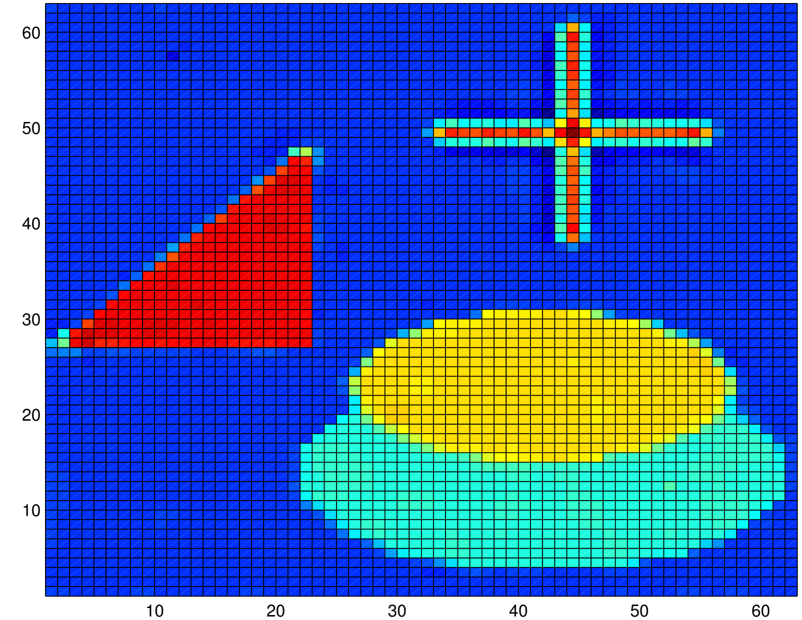
where $TV(f)$ is the discretization of the total variation function

$$\int_0^1 \int_0^1 \left(1 + (\partial_x f)^2 + (\partial_y f)^2\right)\right)^{\frac{1}{2}} \, dx \, dy.$$

Other test problems (5)

Vogel's problem data and result:





Plan
- Introduction
- Problem
- Algorithm
- Taylor iterations
- Iter. structure
- Initial point
- → ***Test problems***
- Some results
- Theory
- Perspectives

The University of Namur

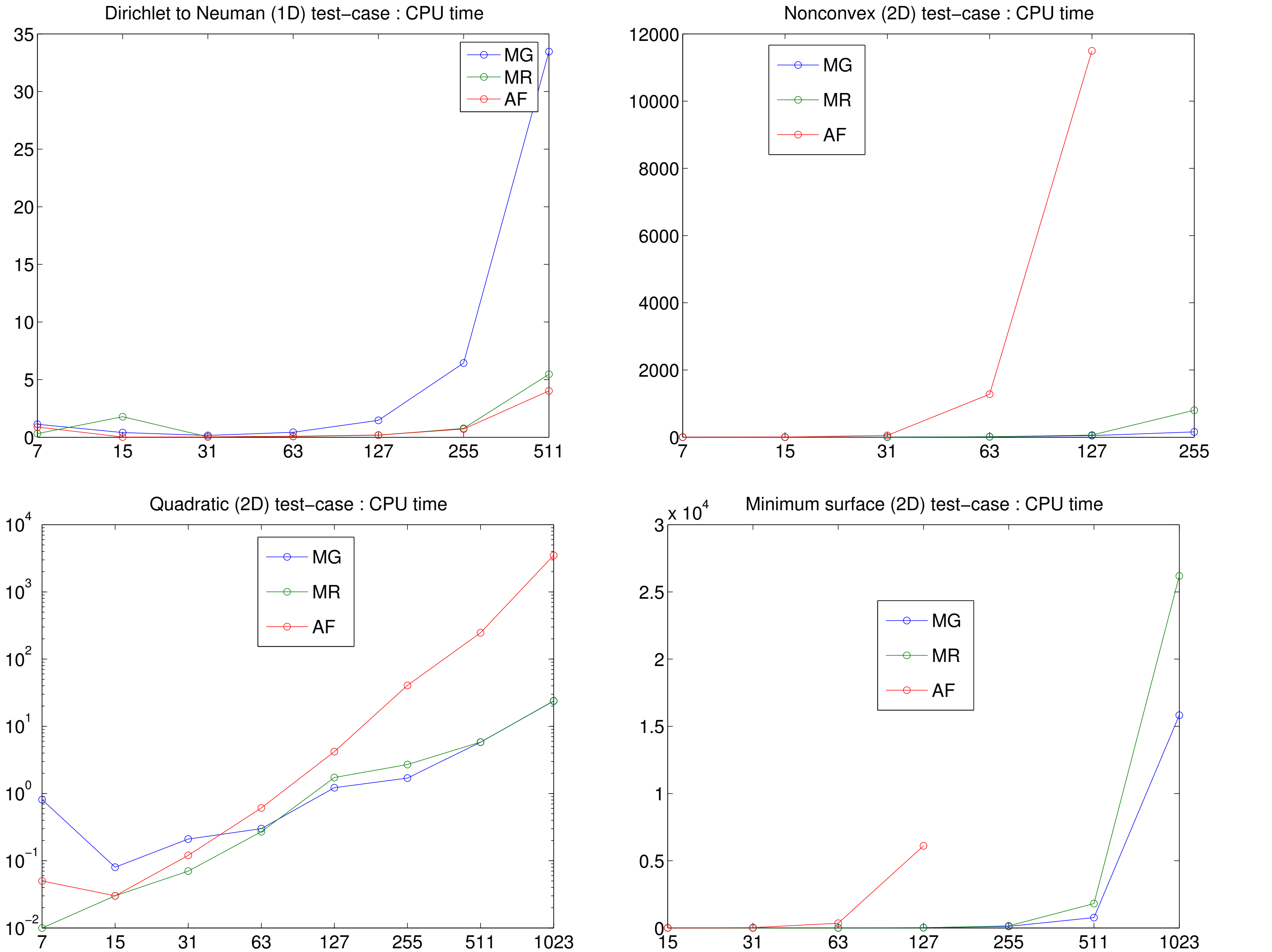
www.fundp.ac.be

A typical run (minimum surface)

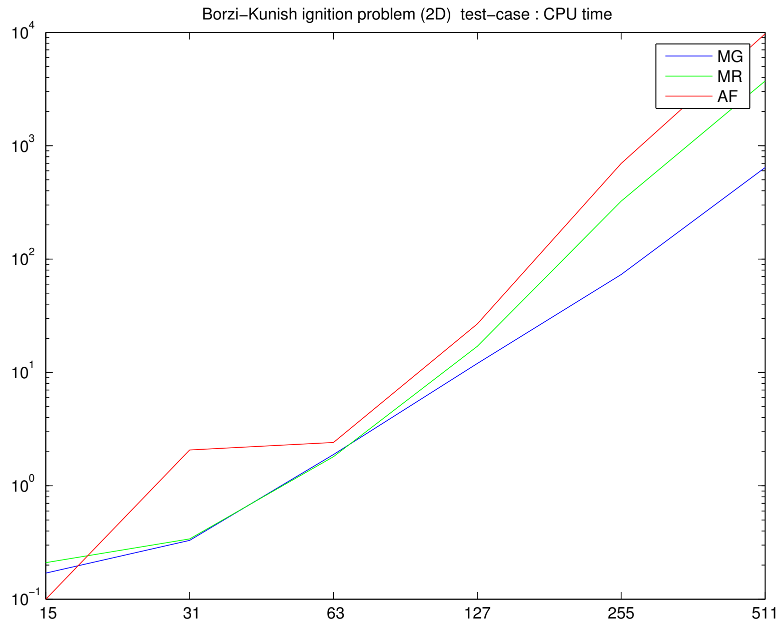
V style, pure quadratic recursion,
2 smoothing cycles, gradient accuracy: 1e-07

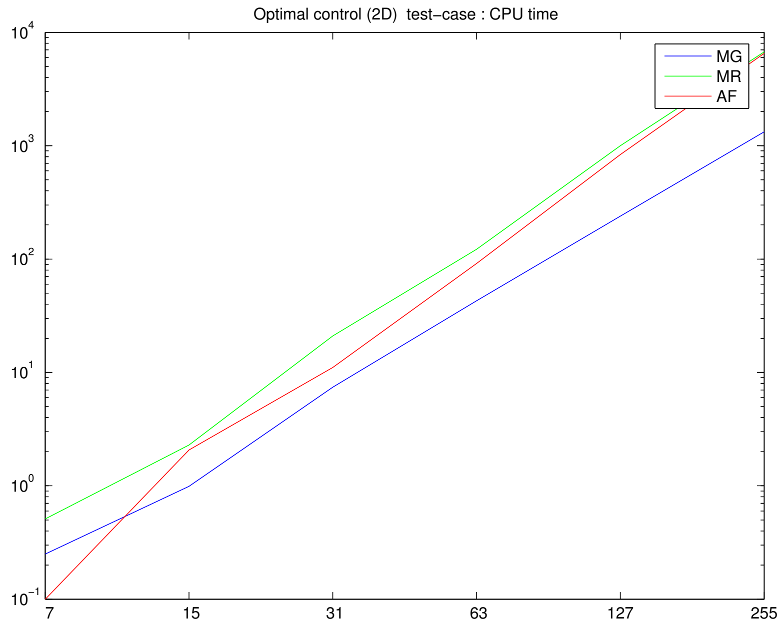
level	3	7	15	31	63	127	255
Tayl. its	10	0	0	0	0	0	0
smooth cyc	0	59	155	232	233	173	70
prolong	0	2	13	25	29	25	20
restric	0	4	26	50	58	50	40
backtrs	0	0	0	0	0	0	1
evals f	12	6	10	18	20	39	72
evals g	6	6	10	18	20	30	42
evals H	6	2	3	6	5	8	16

Problem size and CPU (1)

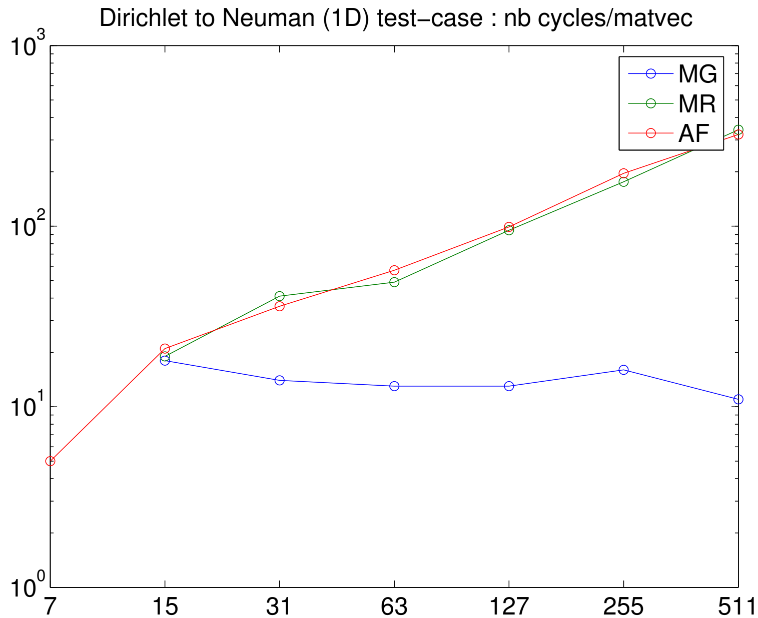


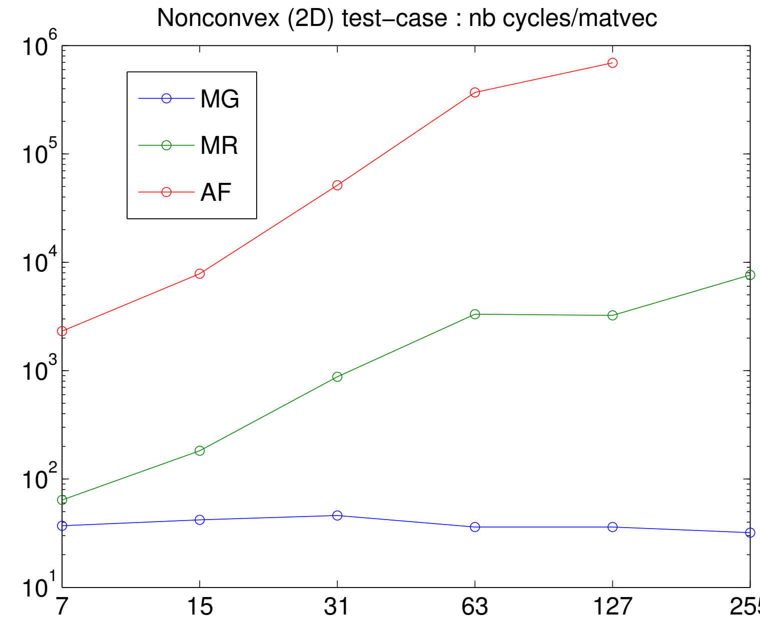
Problem size and CPU (2)

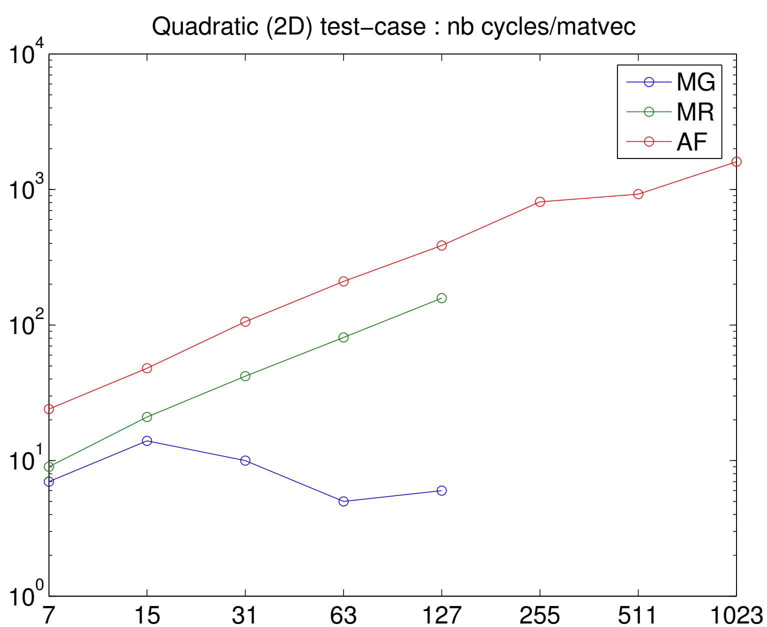


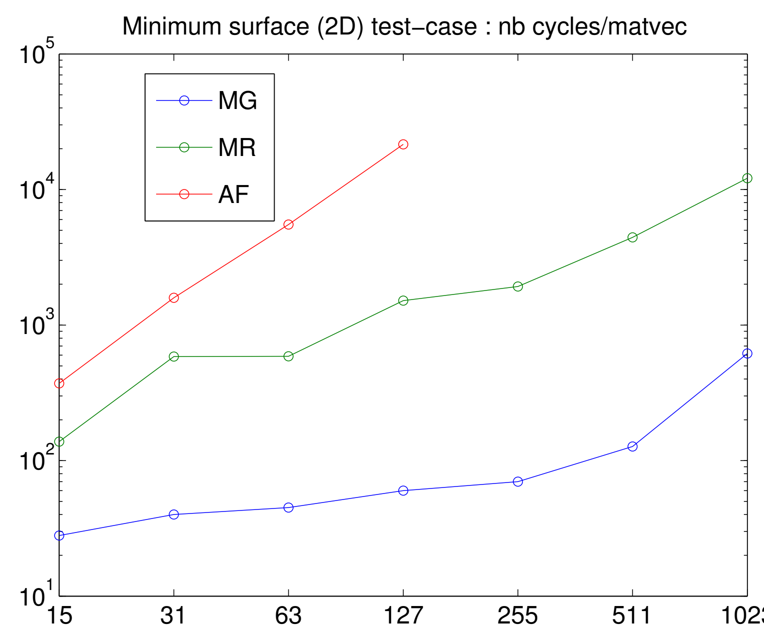


Problem size and linear algebra

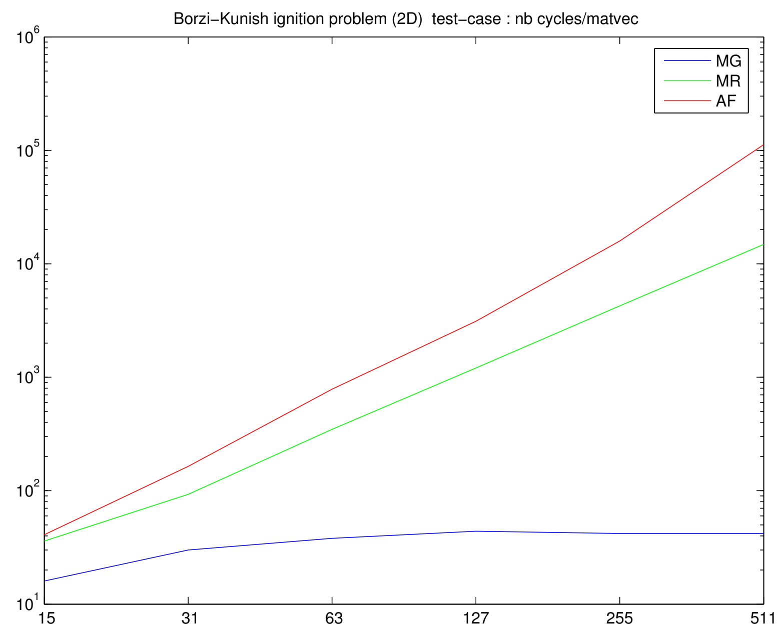


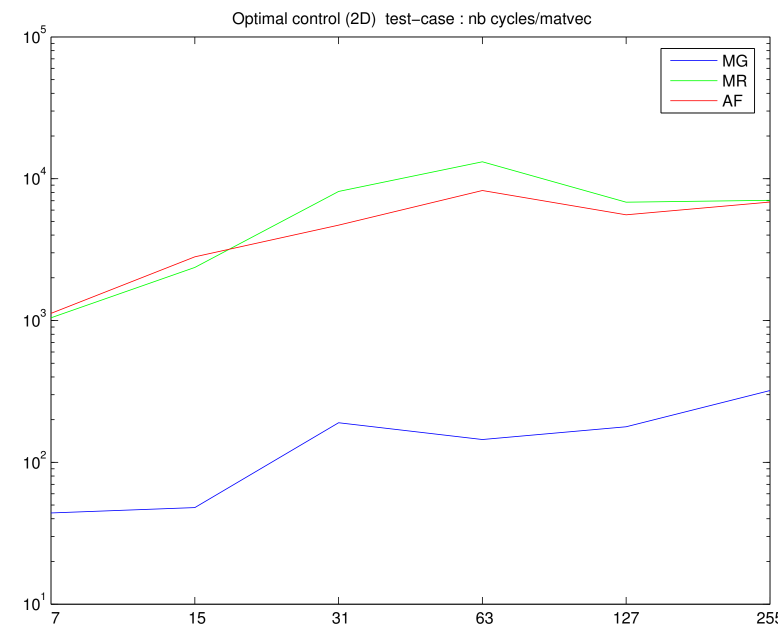




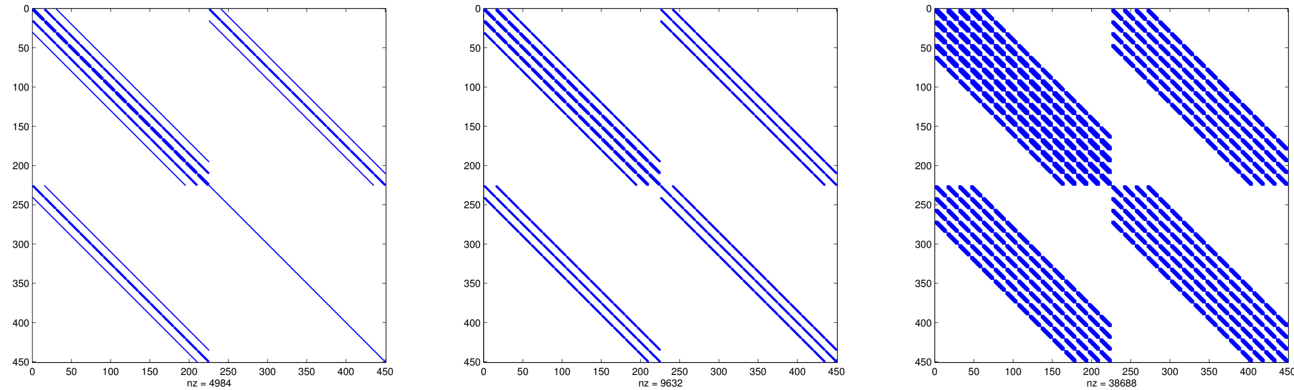


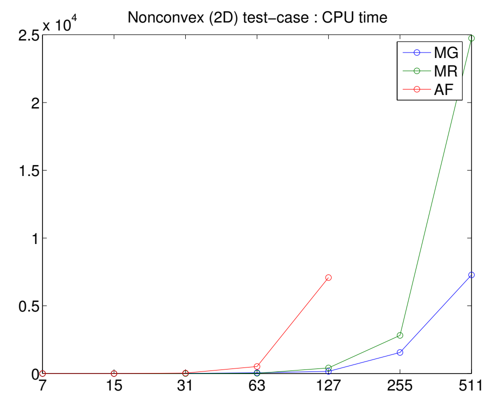
Problem size and linear algebra (2)

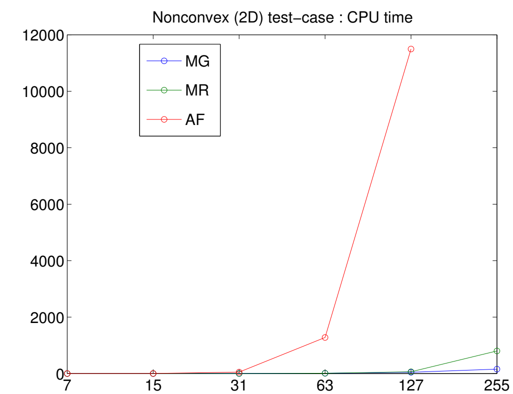


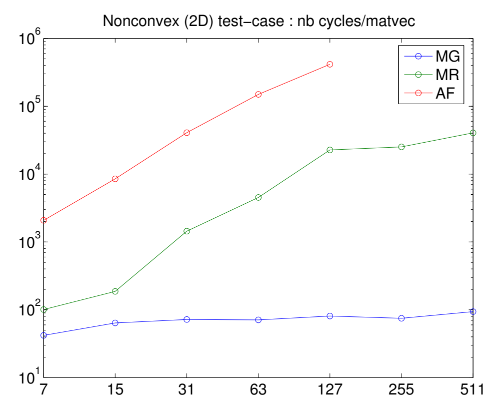


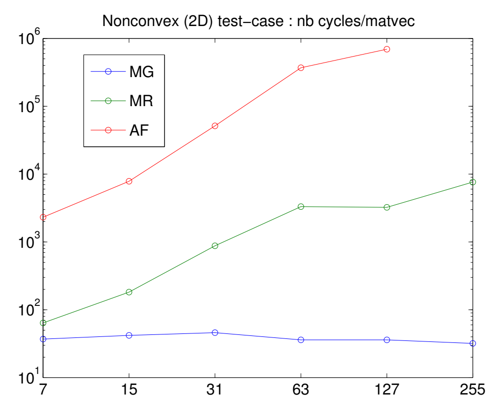
Interpolation and fill-in











Plan

- Introduction
- Problem
- Algorithm
- Taylor iterations
- Iter. structure
- Initial point
- Test problems
- → ***Some results***
- Theory
- Perspectives

Convergence to critical points

Based on trust region technology

- the sufficient decrease argument (imposed in Taylor's iterations)
- a conditional coarsening condition

Main results:

- convergence to first-order critical points at all levels
- (weak) upper bound on the number of iterations to achieve a given accuracy

Convergence to weak minimizers

Difficulties:

- curvature condition too costly to impose a posteriori on recursive iterations: only a subset of iterations can be checked
- Gauss-Seidel only looks at coordinate directions

Main result:

- asymptotic positive curvature on coarsest subspace + prolongation of all coarse coordinate directions

Current conclusions

- more efficient than mesh refinement for large instances
- pure quadratic recursion ($f_i = 0$, Galerkin) very efficient
- interpolation degree crucial
- V cycles most efficient

Perspectives

Encouraging (so far)

- more numerical experiments!
- a (more natural) ℓ_∞ version
- multigrid-type developments:
  (algebraic multilevel, p-multigrids, . . .)
- constrained problems
  (bounds, equalities, general)
- non-monotone (filter) techniques
- . . . and much more!

Thank you for your attention