Multigrid techniques applied in an optimization context

M. Mouffe
CERFACS, Toulouse

joint work with

S. Gratton
CERFACS, Toulouse

Ph. L. Toint
FUNDP, Namur

...and more and more others

29 February 2008

Outline

Outline

Outline

Does it have a sense ?

YES !

Solve a linear system $\Leftrightarrow$ Minimize a quadratic function

$$\min f(x) = \frac{x^T A x}{2} - x^T b + c \Leftrightarrow \nabla_x f(x) = Ax - b = 0 \Leftrightarrow Ax = b$$

Why multigrid

- Solution based on discretization :
  High accuracy $\Rightarrow$ computational cost
- Use of coarse grids :
  1. find a good starting point
  2. solve a subproblem (e.g. the TR subproblem)
- Well-known for solving SPD linear systems resulting of the discretization of a continuous problem
  [W. Briggs, V.E. Henson and S. McCormick, 2000]
- Nonlinear systems
  [W. Hackbuch and A. Reusken, 1989]

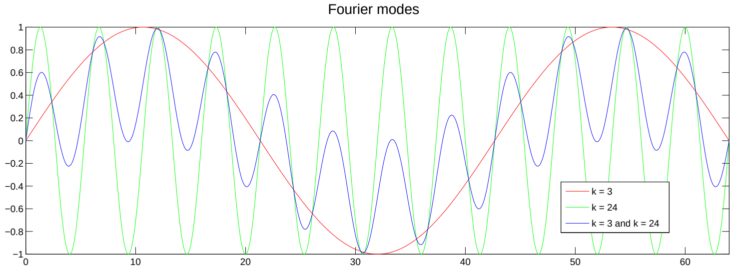
Linear systems

Solve $Ax = b$

$\longrightarrow$ Choice of using an iterative method

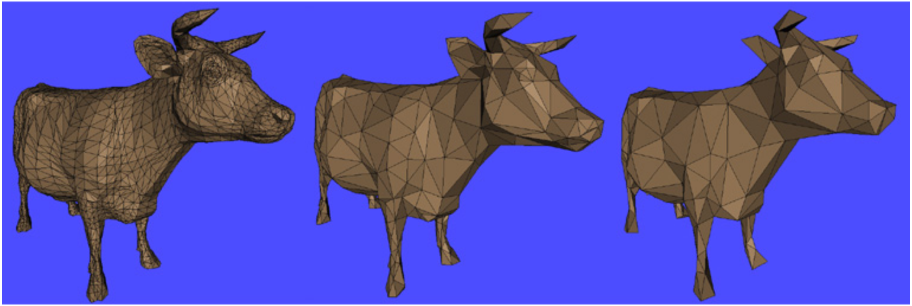
Smoothing/Relaxation methods (Gauss-Seidel)

- Cheap ($\mathcal{O}(n)$)
- Quick in reducing oscillatory components of the error
- Slow in reducing smooth components of the error



Solution :

Use coarser representations



Outline

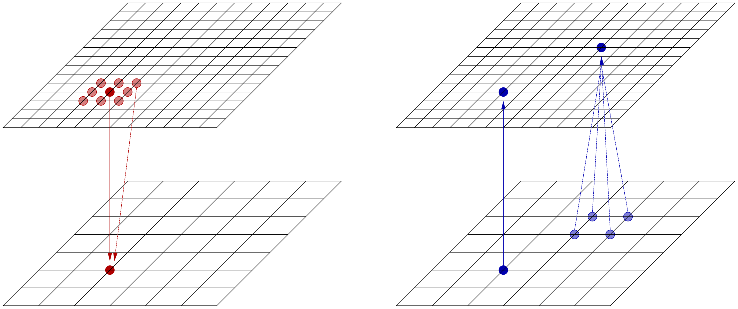
Transfert operators : Geometric case

Restriction

$R_i : \mathbb{R}^{n_i} \longrightarrow \mathbb{R}^{n_{i-1}}$

Prolongation

$P_i : \mathbb{R}^{n_{i-1}} \longrightarrow \mathbb{R}^{n_i}$



$$R_i = \sigma P_i^T$$

Coarse problem definition :

Use the transfert operators

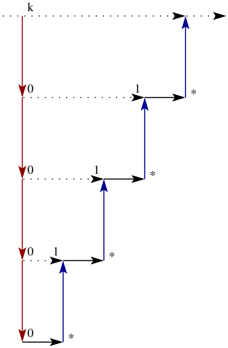
If $A_i x_i = b_i$ is the linear system at level i
Then we define at level $i-1$

$$\boxed{\begin{array}{rcl} A_{i-1} & = & R_i A_i P_i \\ b_{i-1} & = & R_i b_i \end{array}}$$

(Or you already have a coarse definition of A and b)

Mesh refinement : Find a good starting point

- 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



2-levels scheme

Residual equation at level i :

$$\boxed{A_i e_{i,k} = r_{i,k}}$$

where $e_{i,k}$ = error and $r_{i,k}$ = residual at iteration k

Solving a linear system$\Rightarrow$ too expensive

2-levels scheme

Approximate the error using coarse grids

In practice :

1. Compute A_{i-1} and $r_{i-1,k}$
2. Find $e_{i-1,*}$ the solution of the residual equation at level $i-1$
3. Prolongate $e_{i-1,*}$ to define an approximation of the error at level i :

$$\boxed{e_{i,k} = P_i e_{i-1,*}}$$

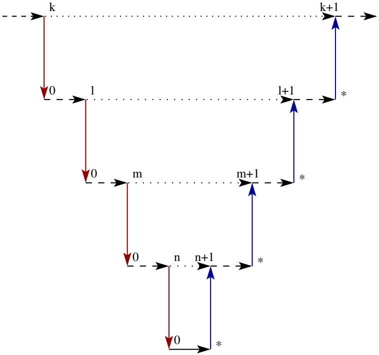
4. Correct your current iterate : $x_{i,k} = x_{i,k} - e_{i,k}$

Combining smoothing and 2-levels scheme

Error behaviour during the multigrid process :

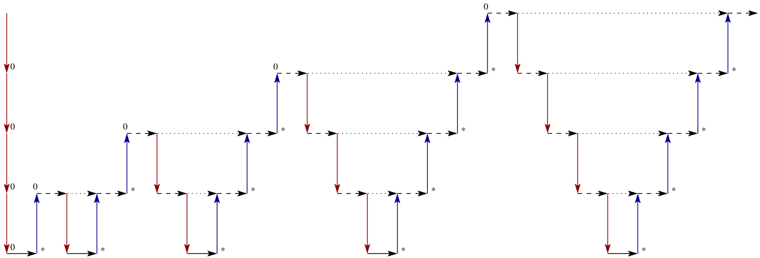
1. Smoothing	Reduces oscillatory error
2. Restriction of the problem	Smooth error appears more oscillatory
3. Smoothing on the coarse grid	Reduces coarse oscillatory error $\Rightarrow$ Reduces smooth fine error
4. Prolongate the solution	Oscillatory error reappears
5. Smoothing again	Reduces this oscillatory error

Recursive use ⇒ V-cycle



Full multilevel scheme

- FMG : Combination of mesh refinement and V-cycles



Outline

Outline

Why trust-region methods

- Newton method : local quadratic convergence
- Trust-region methods : Convergence for all starting point (Global convergence)
- Reduces to the Newton method when close enough to the solution ⇒ Quadratic convergence
- Overview of convergence results and algorithms [A. Conn, N. Gould and Ph. Toint, 2000]

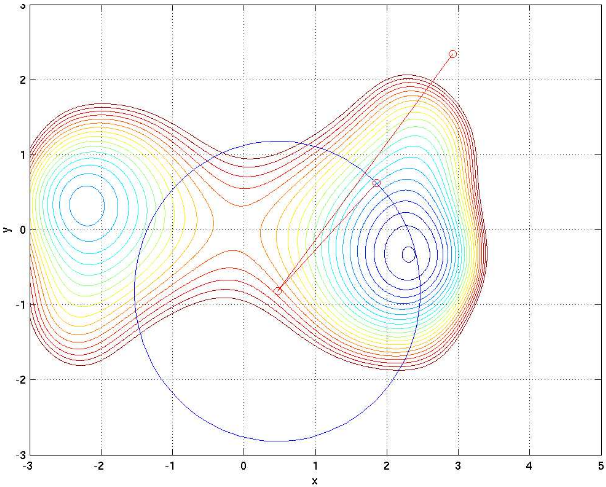
Trust-region mechanism (Bound-constrained optimization)

- Define a model m_k of the objective function f
- Define a trust region where the model is supposed to represent well the objective function
- Compute a step (TR subproblem)
  - inside the TR
  - that sufficiently reduces m_k
  - such that $x_k + s_k \in \{x : l \leq x \leq u\}$
- Step acceptance and TR radius Δ update related to the ratio

$$\frac{f(x_{k+1}) - f(x_k)}{m_k(x_{k+1}) - m_k(x_k)}$$

  - Refuse the step and shrink the TR when the ratio is smaller than a constant
  - Accept the step and possibly enlarge the TR when the ratio is large enough

Trust-region mechanism



Criticality measure and sufficient decrease condition

- Criticality measure : $\chi_k = \left| \min_{\substack{x_k + d \in \mathcal{C} \\ ||d|| \leq 1}} g(x_k)^T d \right|$
  - Unconstrained $\Rightarrow$ Reduces to gradient norm
  - $\chi_k = 0$ at the exact solution
- Stopping criterion : $\chi_k < tol$
- Sufficient decrease condition on the model :

$$m(x_{k+1}) - m(x_k) \geq \kappa \chi_k \min \left[\frac{\chi_k}{1 + ||\nabla^2 f(x_k)||}, \Delta_k, 1 \right]$$

Globally convergent algorithm

Outline

Multigrid ideas in RMTR

- Suppose you have a set of discretizations of the objective function f :

$$\{f_i\}_{i=0}^{r}$$

  with $f_r = f$.

- Transfert operators :

$$\begin{array}{llcll} R_i : & \mathbb{R}^{n_i} & \rightarrow & \mathbb{R}^{n_{i-1}} & \text{Restriction} \\ P_i : & \mathbb{R}^{n_{i-1}} & \rightarrow & \mathbb{R}^{n_i} & \text{Prolongation} \end{array}$$

- Coarse model : Need to be first order coherent

$$\Leftrightarrow$$

"tau correction" in multigrid

Smoothing

In multigrid : Smoothing = Solving the equations of the linear system one by one

In optimization :

- Smoothing = Solving the minimization problem along the coordinate axes j.
- Bound-constrained unidirectional problem (Trust-region and possibly original bounds constraints)
- The final step is defined by $s = \sum_j s_j$

Smoothing ⇒ Sufficient decrease

For unconstrained optimization

If the minimization begins in the direction $\text{argmax}_j |g_j|$

Convergence theory available

[S. Gratton, A. Sartenaer and Ph. Toint, 2005]

For bound-constrained optimization

If the minimization begins in the direction $\text{argmax}_j g_j^T d_j$

where d is defined by $\left|\text{argmin}_{\substack{x_k + d \in \mathcal{C} \\ \|d\| \leq 1}} g(x_k)^T d\right|$

Coarse step

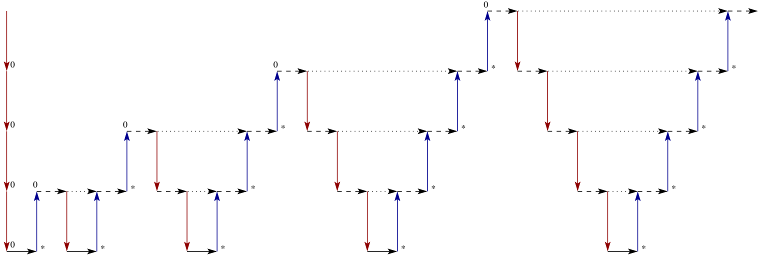
Step computation on a coarse level :

- Define $x_{i-1,0} = R_i x_{i,k}$
- Find a coarse step s_{i-1} (e.g. using smoothing)
  - inside a coarse version of the TR
  - inside a coarse version of the bounds, and
  - that reduces sufficiently a coarse model f_{i-1} of the objective function f_i
- Use P_i to obtain a fine step s_i by $s_i = P_i s_{i-1}$
- Coarse sufficient decrease $\Rightarrow$ fine sufficient decrease

Full multigrid (FMG)

As in multigrid methods :

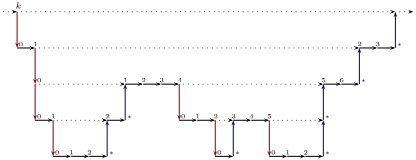
- Mesh refinement
- V-cycles



Outline

Free cycles

Solve at a coarse level i until $\chi_i < \varepsilon_i$



Example of recursion with 5 levels ($r = 4$)

Descent condition

Use coarse levels only if

$$\boxed{\chi_{i-1,0} \geq \kappa_{\chi} \chi_{i,k}}$$

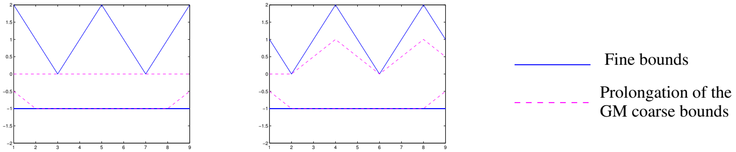
not worth working on coarser level
if the problem is already solved there

Coarse problem definition

- Coarse TR : Restriction of the fine TR using R_i
- Coarse bounds : Gelman & Mandel's definition

$$[l_{r-1}]_j = [R_r x_{r,k}]_j + \max_{\substack{t=1,\ldots,n_r \\ [P_r]_{tj}>0}} [l - x_{r,k}]_t$$

$$[u_{r-1}]_j = [R_r x_{r,k}]_j + \min_{\substack{t=1,\ldots,n_r \\ [P_r]_{tj}>0}} [u - x_{r,k}]_t$$



Nonlinear multigrid ?

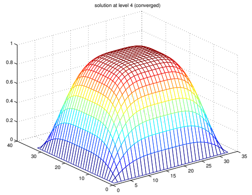
Multigrid : Solving a 1st order Taylor approximation
(e.g. a linear system)

$\Rightarrow$

RMTR is equivalent if a 2nd order Taylor model is used in the trust-region method (e.g. quadratic minimization problem)

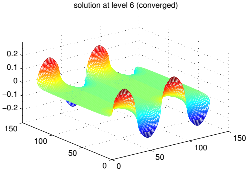
Outline

Some treated problems



- Quadratic minimization :

  $\min_u u^T A u - 2u^T b \Leftrightarrow Au = b$



- Minimal surface :

  $\min_u \int\int \sqrt{1 + u_x^2 + u_y^2} dxdy$

Quadratic minimization

	Mesh Refinement	$RMTR_\infty$
Nb iter	365	6
Nb eval f	2	11
Nb eval g	2	11
Nb eval H	1	1

Minimal surface with obstacle

	Mesh Refinement	RMTR_∞
Nb iter	1410	111
Nb eval f	95	182
Nb eval g	85	177
Nb eval H	58	29

Outline

Conclusion

RMTR :

- Optimization algorithm inspired by multigrid ideas
- Adapts multigrid techniques
- Proved globally convergent
- Adapted to bound-constrained problems
- Very efficient on discretized problems

Thank you !