Multilevel Optimization

Annick Sartenaer

University of Namur, Belgium

SIAM Conference on Optimization, Boston, May 2008

The problem

$$\min_{x \in \mathbb{R}^n} f(x)$$

- $f : \mathbb{R}^n \rightarrow \mathbb{R}$ nonlinear, $\in \mathcal{C}^2$ and bounded below
- No convexity assumption
- Results from the discretization of some infinite-dimensional problem on a relatively fine grid for instance (n large)

$\longrightarrow$ Iterative search of a first-order critical point x_* (s.t. $\nabla f(x_*) = 0$)

Newton's method

$$x_{k+1} = x_k + s_k^N \quad \text{with} \quad \nabla^2 f(x_k)\, s_k^N = -\nabla f(x_k)$$

- Fast convergence (quadratic) to a local minimizer x_* of f
- If x_0 sufficiently close to x_*

$\longrightarrow$ Requires a globalization technique in order to:

– Ensure convergence of the iterates from every starting point

– Take account of the nonconvexity when far from a local min.

Line search — Trust region — Adaptive regularization

Hierarchy of problem descriptions

Assume now that a hierarchy of problem descriptions is available, linked by known operators

Finest problem description

Restriction $\downarrow R$ $\quad$ $P \uparrow$ Prolongation

Fine problem description

Restriction $\downarrow R$ $\quad$ $P \uparrow$ Prolongation

$\cdots$

Restriction $\downarrow R$ $\quad$ $P \uparrow$ Prolongation

Coarse problem description

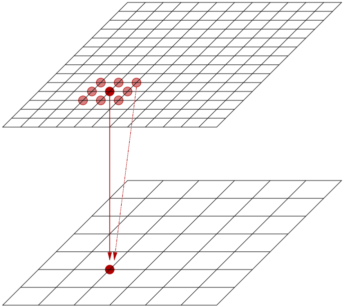
Restriction $\downarrow R$ $\quad$ $P \uparrow$ Prolongation

Coarsest problem description

Grid transfer operators

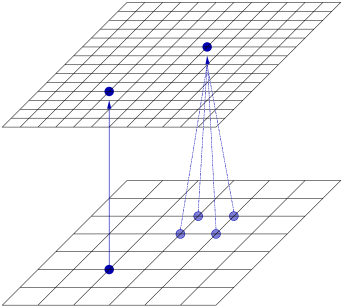
Restriction

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



Prolongation

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



$$\boxed{R_i = \sigma P_i^T}$$

Sources for such problems

- Parameter estimation in
  - discretized ODEs
  - discretized PDEs
- Optimal control problems
- Variational problems (minimum surface problem)
- Optimal surface design (shape optimization)
- Data assimilation in weather forecast (different levels of physics in the models)

The minimum surface problem

$$\boxed{\min_{v} \int_0^1 \int_0^1 \left(1 + (\partial_x v)^2 + (\partial_y v)^2\right)^{\frac{1}{2}} \, dx \, dy}$$

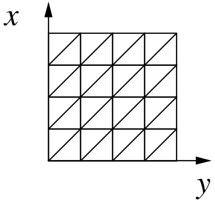
with the boundary conditions:

$$\begin{cases} f(x), & y = 0, \quad 0 \le x \le 1 \\ 0, & x = 0, \quad 0 \le y \le 1 \\ f(x), & y = 1, \quad 0 \le x \le 1 \\ 0, & x = 1, \quad 0 \le y \le 1 \end{cases}$$

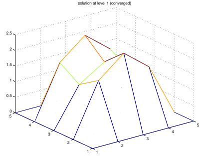
where

$$f(x) = x * (1 - x)$$

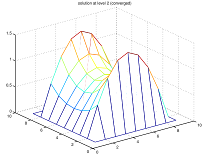
$\longrightarrow$ Discretization using a finite element basis



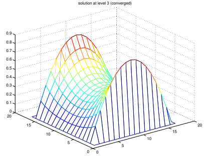
The solution at different levels



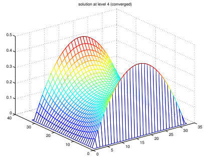
$n = 3^2 = 9$



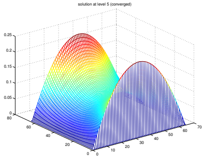
$n = 7^2 = 49$



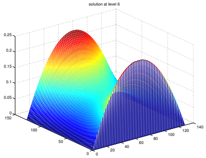
$n = 15^2 = 225$



$n = 31^2 = 961$



$n = 63^2 = 3969$



$n = 127^2 = 16129$

The main issue

Hierarchy of problem descriptions

globalization technique

$\searrow \qquad \swarrow$

Efficiency – Robustness

$\Downarrow$

Illustration within a trust-region framework

(Unconstrained case)

Past and recent developments

Line-search

- Fisher (1998), Frese-Bouman-Sauer (1999), Nash (2000)
- Lewis-Nash (2000, 2002, to appear)
- Oh-Milstein-Bouman-Webb (2003)
- Wen-Goldfarb (2007, report 2008)
- Gratton-Toint (report 2007)

Trust-region

- Gratton-Sartenaer-Toint (to appear in SIOPT)
- Gratton-Mouffe-Toint-Weber Mendonça (to appear in IMAJNA)
- Mouffe-Gratton-Sartenaer-Toint-Tomanos (in preparation)
- Toint-Tomanos-Weber Mendonça (report 2007)

Adaptive regularization

- Toint-Tomanos (in preparation)

A very active field

- Large Scale Optimization and PDE-Based Problems (MS2, MS12, MS42, MS52)
- Multigrid/Multilevel Optimization Methods and Their Applications (MS62)
- Numerical Treatment of PDE Constrained Optimization Problems:
  - A: Numerical Analysis (MS3, MS13)
  - B: Algorithms (MS23, MS33)
  - C: Applications (MS53, MS63, MS73)
- Optimization with PDE Constraints (CP3)

Too late! To come

Outline

1. Trust-region methods for beginners
2. Multigrid for beginners
3. RMTR (a Recursive Multilevel Trust-Region Method)
   - Theoretical aspects
   - Practical aspects
4. Some numerical flavor

Trust-region philosophy

At iteration k (until convergence):

- Choose a local model m_k of f around x_k (Taylor's model)
- Compute a trial step s_k that suff. reduces m_k in a trust region:

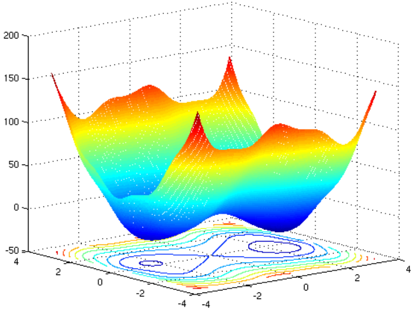
$$\begin{cases} \text{(approx.) minimize}_{s \in \mathbb{R}^n} & m_k(x_k + s) \\ \text{subject to} & \|s\| \leq \Delta_k \end{cases}$$

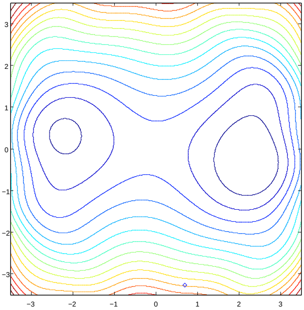
- Evaluate $f(x_k + s_k)$
- If achieved decrease (Δf) $\approx$ predicted decrease (Δm_k), then
  - accept the trial point $(x_{k+1} = x_k + s_k)$
  - possibly enlarge the trust region $(\Delta_k \nearrow)$

  else

  - keep the current point $(x_{k+1} = x_k)$
  - shrink the trust region $(\Delta_k \searrow)$

minimize : $f(\alpha, \beta) = -10\alpha^2 + 10\beta^2 + 4\sin(\alpha\beta) - 2\alpha + \alpha^4$

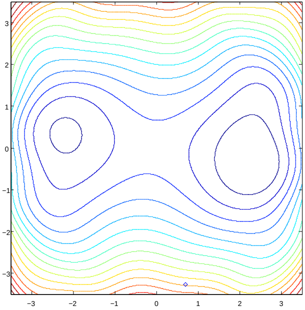




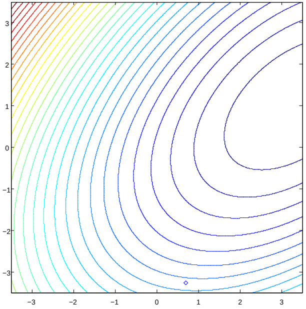
Two local minima: $(-2.20, 0.32)$ and $(2.30, -0.34)$

$$x_0 = (0.71, -3.27) \quad \text{and} \quad f(x_0) = 97.630$$

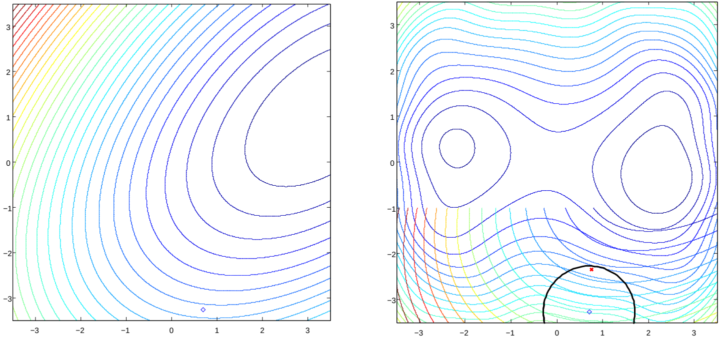
Contours of f



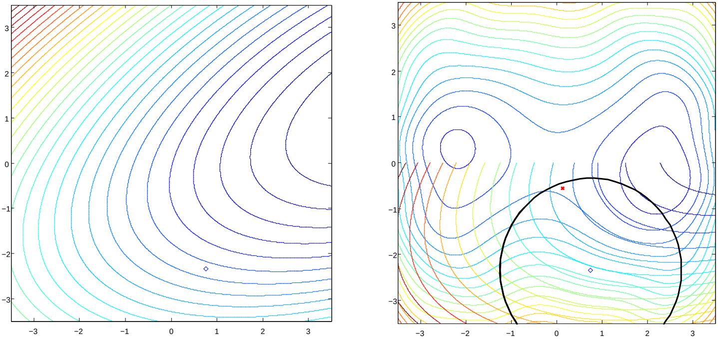
Contours of m_0 around x_0
(quadratic model)



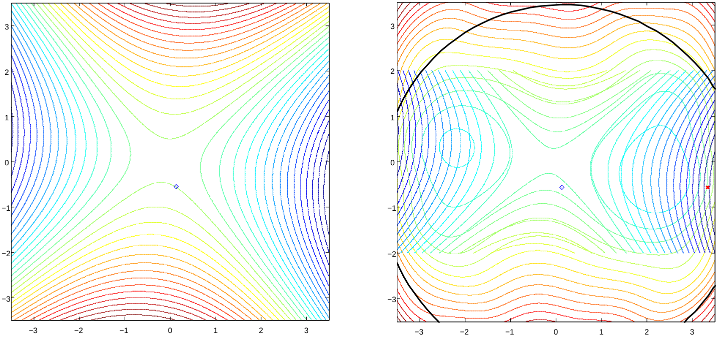
k	Δ_k	s_k	$f(x_k + s_k)$	$\Delta f / \Delta m_k$	x_{k+1}
0	1	$(0.05, 0.93)$	43.742	0.998	$x_0 + s_0$



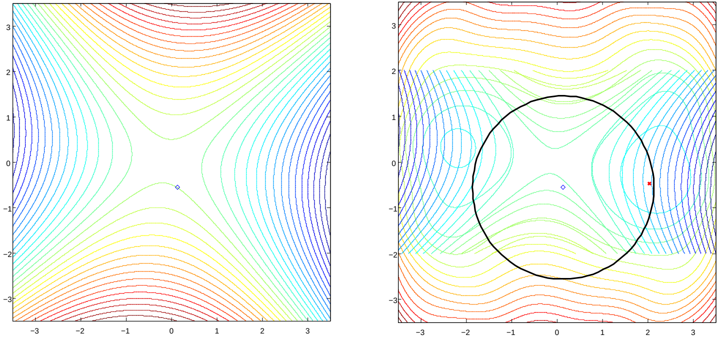
k	Δ_k	s_k	$f(x_k + s_k)$	$\Delta f / \Delta m_k$	x_{k+1}
0	**1**	$(0.05, 0.93)$	43.742	0.998	$x_0 + s_0$
1	**2**	$(-0.62, 1.78)$	2.306	1.354	$x_1 + s_1$



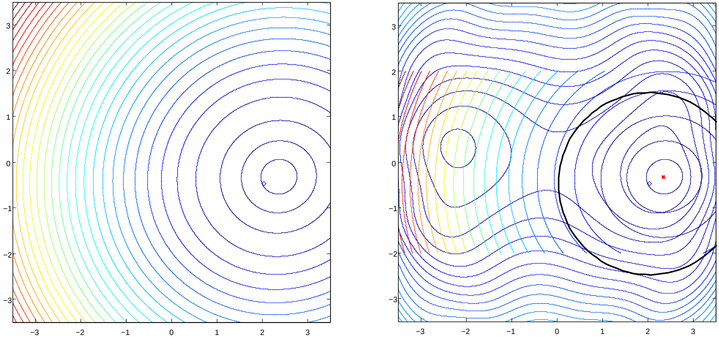
k	Δ_k	s_k	$f(x_k + s_k)$	$\Delta f / \Delta m_k$	x_{k+1}
0	**1**	$(0.05, 0.93)$	43.742	0.998	$x_0 + s_0$
1	**2**	$(-0.62, 1.78)$	2.306	1.354	$x_1 + s_1$
2	**4**	$(3.21, 0.00)$	6.295	-0.004	x_2



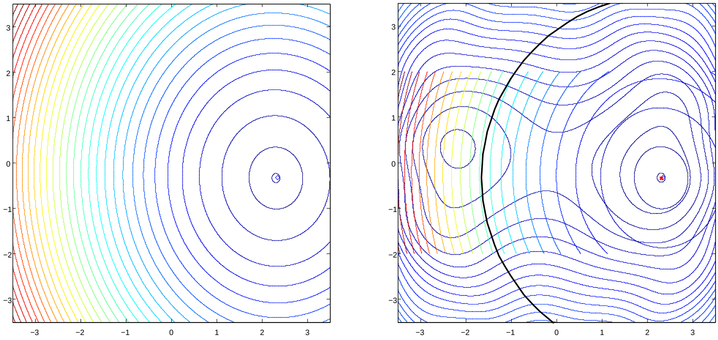
k	Δ_k	s_k	$f(x_k + s_k)$	$\Delta f/\Delta m_k$	x_{k+1}
0	**1**	$(0.05, 0.93)$	43.742	0.998	$x_0 + s_0$
1	**2**	$(-0.62, 1.78)$	2.306	1.354	$x_1 + s_1$
2	**4**	$(3.21, 0.00)$	6.295	-0.004	x_2
3	**2**	$(1.90, 0.08)$	-29.392	0.649	$x_2 + s_2$



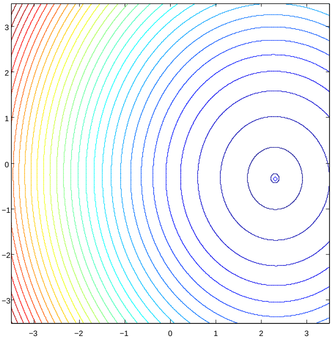
k	Δ_k	s_k	$f(x_k + s_k)$	$\Delta f / \Delta m_k$	x_{k+1}
0	**1**	$(0.05, 0.93)$	43.742	0.998	$x_0 + s_0$
1	**2**	$(-0.62, 1.78)$	2.306	1.354	$x_1 + s_1$
2	**4**	$(3.21, 0.00)$	6.295	-0.004	x_2
3	**2**	$(1.90, 0.08)$	-29.392	0.649	$x_2 + s_2$
4	**2**	$(0.32, 0.15)$	-31.131	0.857	$x_3 + s_3$

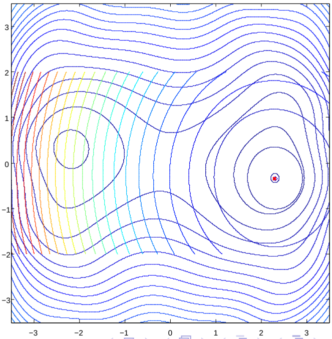


k	Δ_k	s_k	$f(x_k + s_k)$	$\Delta f / \Delta m_k$	x_{k+1}
0	**1**	$(0.05, 0.93)$	43.742	0.998	$x_0 + s_0$
1	**2**	$(-0.62, 1.78)$	2.306	1.354	$x_1 + s_1$
2	**4**	$(3.21, 0.00)$	6.295	-0.004	x_2
3	**2**	$(1.90, 0.08)$	-29.392	0.649	$x_2 + s_2$
4	**2**	$(0.32, 0.15)$	-31.131	0.857	$x_3 + s_3$
5	**4**	$(-0.03, -0.02)$	-31.176	1.009	$x_4 + s_4$

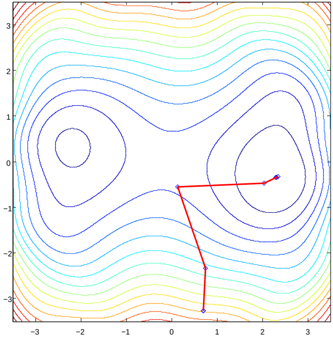


k	Δ_k	s_k	$f(x_k + s_k)$	$\Delta f / \Delta m_k$	x_{k+1}
0	**1**	$(0.05, 0.93)$	43.742	0.998	$x_0 + s_0$
1	**2**	$(-0.62, 1.78)$	2.306	1.354	$x_1 + s_1$
2	**4**	$(3.21, 0.00)$	6.295	-0.004	x_2
3	**2**	$(1.90, 0.08)$	-29.392	0.649	$x_2 + s_2$
4	**2**	$(0.32, 0.15)$	-31.131	0.857	$x_3 + s_3$
5	**4**	$(-0.03, -0.02)$	-31.176	1.009	$x_4 + s_4$
6	**8**	$(-0.02, 0.00)$	-31.179	1.013	$x_5 + s_5$

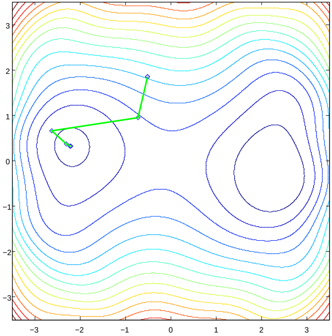




Path of iterates:



From another x_0:



What makes it work ?

$$\begin{cases} \text{(approx.) minimize}_{s\in\mathbb{R}^n} & m_k(x_k+s) \\ \text{subject to} & \|s\| \le \Delta_k \end{cases}$$

$(g_k = \nabla m_k(x_k) = \nabla f(x_k))$

$$\downarrow$$

$$\begin{cases} \text{minimize}_{\,t\geq 0} & m_k(x_k - t g_k) \\ \text{subject to} & t\|g_k\| \le \Delta_k \end{cases}$$

$$\downarrow$$

Model decrease at the Cauchy point x_k^C

(Best decrease of the model within the trust region along the steepest descent direction $-g_k$)

$$\boxed{m_k(x_k) - m_k(x_k^C) \geq \frac{1}{2}\|g_k\| \min\left[\frac{\|g_k\|}{\beta_k}, \Delta_k\right]}$$

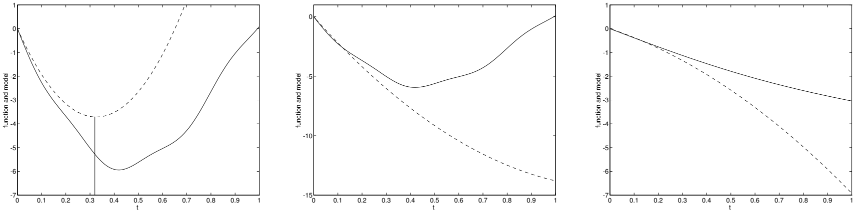


Illustration from [Conn, Gould, Toint, 2000]

positive curvature and minimum inside	$m_k(x_k) - m_k(x_k^C) \geq \frac{1}{2}\frac{\|g_k\|^2}{\beta_k}$
positive curvature and minimum outside	$m_k(x_k) - m_k(x_k^C) \geq \frac{1}{2}\|g_k\|\Delta_k$
negative curvature	$m_k(x_k) - m_k(x_k^C) \geq \|g_k\|\Delta_k$

$\beta_k =$ upper bound on the curvature of m_k

First-order convergence

Sufficient decrease condition:

$$\boxed{m_k(x_k) - m_k(x_k + s_k) \geq \kappa \left(m_k(x_k) - m_k(x_k^C)\right)}$$

$$\Downarrow$$

$$\boxed{\lim_{k\to\infty} \|\nabla f(x_k)\| = 0}$$

What makes it fast ?

$$m_k(x_k + s) = f(x_k) + g_k^T s + \tfrac{1}{2} s^T H_k s$$

where $g_k = \nabla f(x_k)$ and $H_k \approx \nabla^2 f(x_k)$ (possibly indefinite)

Any global minimizer s^* of

$$\begin{cases} \text{minimize}_{s \in \mathbb{R}^n} & m_k(x_k + s) \\ \text{subject to} & \|s\|_2 \leq \Delta_k \end{cases}$$

satisfies:

$$(H_k + \lambda^* I)\, s^* = -g_k$$

where $H_k + \lambda^* I$ is pos. (semi)def., $\lambda^* \geq 0$ and $\lambda^*(\|s^*\|_2 - \Delta_k) = 0$

Exact solution: search for λ^* (Moré-Sorensen)

For Δ_k fixed, find $\lambda \geq \max\{0, -\lambda_{\min}(H_k)\}$ such that:

- $H_k + \lambda I$ is positive semidefinite
- $\boxed{s(\lambda) = -(H_k + \lambda I)^{-1}\, g_k}$ satisfies $\begin{cases} \|s(\lambda)\|_2 \leq \Delta_k & \text{for } \lambda = 0 \\ \|s(\lambda)\|_2 - \Delta_k = 0 & \end{cases}$

by applying a safeguarded Newton's method to the secular equation:

$$\boxed{\phi(\lambda) \stackrel{\text{def}}{=} \frac{1}{\|s(\lambda)\|_2} - \frac{1}{\Delta_k} = 0}$$

Dominating cost: $s(\lambda)$ (a small number of Cholesky factorizations)

Inexact solution: Krylov method (Steihaug-Toint)

- Adapt the (preconditioned) conjugate gradient method:
  - iterative method (n iterations) that generates a sequence $\{p_j\}$ of mutually conjugate directions with respect to H_k:

$$p_j^T H_k p_i = 0 \quad i \neq j$$

  - along which $m_k(x_k + s)$ is exactly minimized

- for the solution of the trust-region subproblem:

$$\begin{cases} \text{(approx) } \min_{s \in \mathbb{R}^n} & m_k(x_k + s) \\ \text{subject to} & \|s\|_2 \leq \Delta_k \end{cases}$$

Start from the Cauchy point x_k^C (that is, with $p_0 = -g_k$)

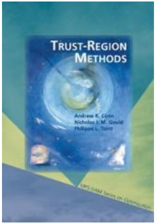
- in order to ensure a further reduction in the model m_k

Terminate

- when an approximate minimizer is found (Stop)
- when the trust-region boundary is passed (Stop at the boundary)
- when a direction of negative curvature is encountered (move to the boundary and Stop)

For instance:
- The Steihaug-Toint algorithm
- The Generalized Lanczos Trust-Region algorithm (GLTR)

Book on trust-region methods



Trust-region methods [Conn, Gould, Toint, 2000]

On the side of multigrid methods

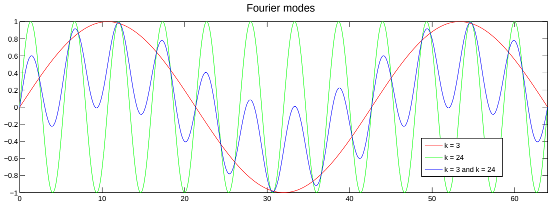
Consider the linear system (discrete Poisson equation, for instance):

$$\boxed{Ax = b} \quad \rightsquigarrow \quad \boxed{Ae = r} \qquad \text{(residual equation)}$$

where

- $e = x_* - \tilde{x}$ (error)
- $r = b - A\tilde{x}$ (residual)
- x_* (exact solution)
- $\tilde{x}$ (approximation)

Expansion of e in Fourier modes shows high (oscillatory) and low (smooth) frequency components:



Relaxation methods

Basic iterative methods:

- correct the i^{th} component of x_k in the order $1, 2, \ldots, n$
- to annihilate the i^{th} component of r_k

Jacobi

$$[x_{k+1}]_i = \frac{1}{a_{ii}} \left(- \sum_{j=1,\ j\neq i}^{n} a_{ij} [x_k]_i + [b]_i \right)$$

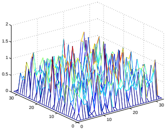
Gauss-Seidel

$$[x_{k+1}]_i = \frac{1}{a_{ii}} \left(- \sum_{j=1}^{i-1} a_{ij} [x_{k+1}]_i - \sum_{j=i+1}^{n} a_{ij} [x_k]_i + [b]_i \right)$$

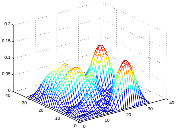
$\longrightarrow$ Solve the equations of the linear system one by one

Smoothing effect

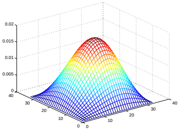
Very effective methods at “smoothing”, i.e., eliminating the high-frequency (oscillatory) components of the error:



error of
initial guess



error after 10
GS iterations



error after 100
GS iterations

But they leave the low-frequency (smooth) components relatively unchanged

Assume now (two levels):

- A fine grid (f) description $\quad Ae = r \quad \longrightarrow \quad A^f e^f = r^f$
- A coarse grid (c) description $\quad A^c e^c = r^c$
- Linked by transfer operators $\quad A^c = RA^f P, \quad e^c = Re^f, \quad r^c = Rr^f$

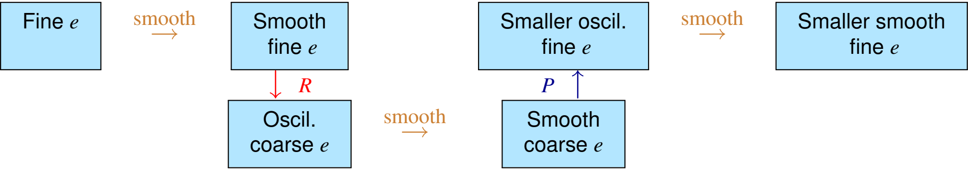
Coarse grid principle

Smooth error modes on a fine grid
"look less smooth" on a coarse grid

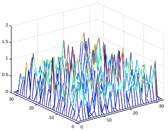
$\longrightarrow$ When relaxation begins to stall at the finer level:

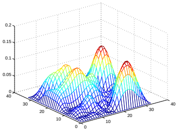
- Move to the coarser grid where the smooth error modes appear more oscillatory
- Apply a relaxation at the coarser level:
  - more efficient
  - substantially less expensive

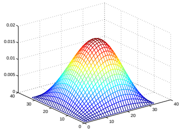
Two-grid correction scheme



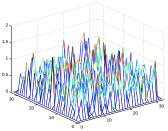
Smoothing on fine grid only:

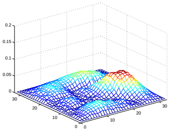


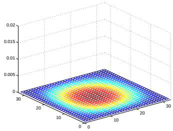




Two-grid correction scheme:

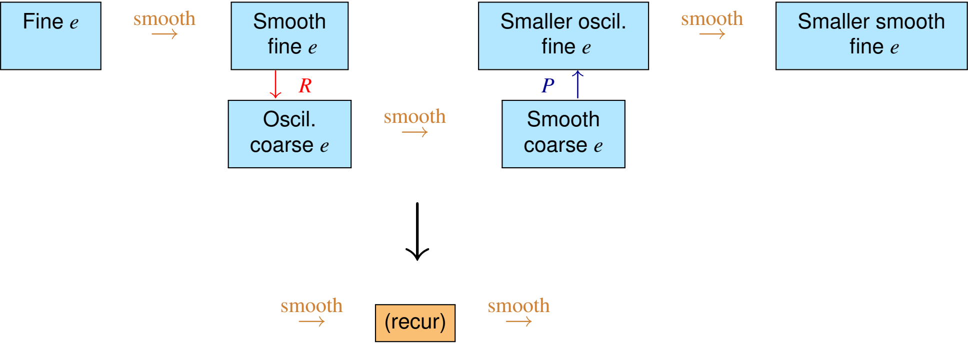






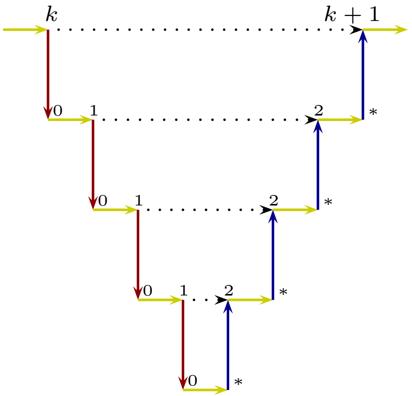
$k = 0$ $\qquad$ $k = 10$ $\qquad$ $k = 100$

Multigrid scheme



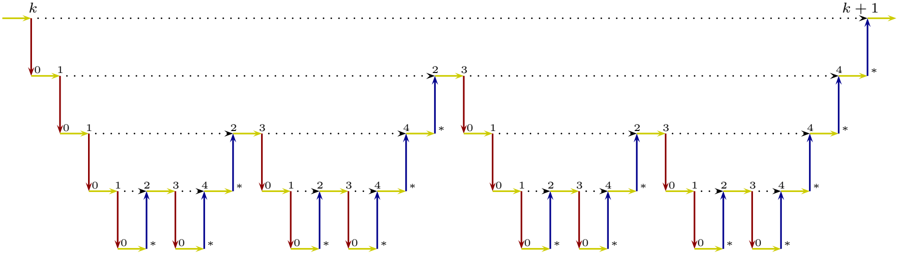
Recursive use to annihilate oscillatory error level by level $(\mathcal{O}(n))$

V-cycle



Smoothing

W-cycle



Smoothing

Mesh Refinement

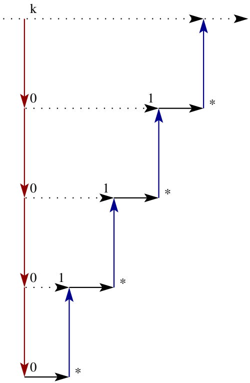
- Solve the problem on the coarsest level

  $\Rightarrow$ Good starting point for the next fine level

- Do the same on each level

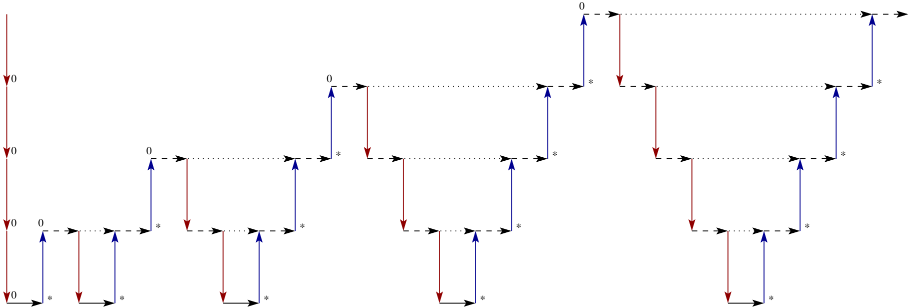
  $\Rightarrow$ Good starting point for the finest level

- Finally solve the problem on the finest level

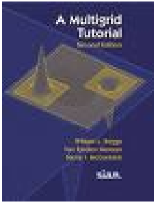


Full Multigrid

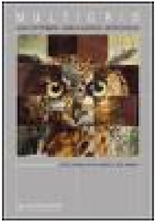
Combination of Mesh Refinement and V-cycles



Books on multigrid



A Multigrid Tutorial [Briggs, Henson, McCormick, 2000]



Multigrid [Trottenberg, Oosterlee, Schüller, 2001]

Back to our main issue

Hierarchy of problem descriptions

Trust-region technique

$\searrow \qquad \swarrow$

Efficiency – Robustness

$\Downarrow$

Multilevel **optimization** method

Note: Multilevel Moré-Sorensen algorithm: $(H_k + \lambda I)\, s = -g_k$

[Toint, Tomanos, Weber Mendonça, report 2007]

The framework

Assume that we have:

- A hierarchy of problem descriptions of f:

$$\{f_i\}_{i=0}^{r} \qquad \text{with} \qquad f_r(x) = f(x)$$

- Transfer operators, for $i = 1, \ldots, r$:
  - $R_i \colon \mathbb{R}^{n_i} \to \mathbb{R}^{n_{i-1}}$ (the restriction)
  - $P_i \colon \mathbb{R}^{n_{i-1}} \to \mathbb{R}^{n_i}$ (the prolongation)

Terminology: a particular i is referred to as a level

The idea

$\min_{x \in \mathbb{R}^n} f_r(x) = f(x)$ $\rightarrow$ at x_k: minimize Taylor's model of f_r around x_k in the trust region of radius Δ_k

$\downarrow$ or (whenever suitable and desirable)

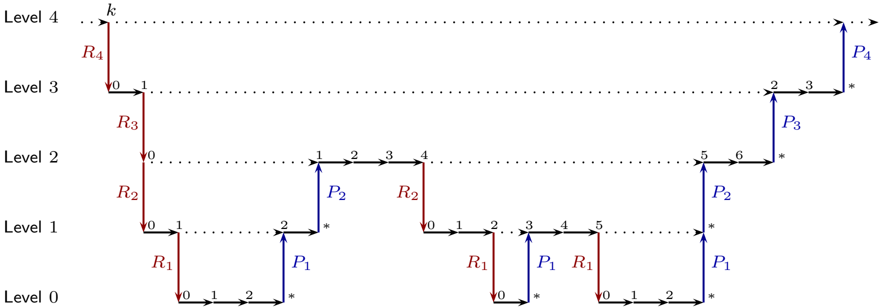
at x_k: compute $\nabla f_r(x_k)$ (possibly H_k) trial step s_k

Restriction $\downarrow$ R P $\uparrow$ Prolongation

use f_{r-1} to construct a **coarse local** model of f_r and minimize it within the trust region of radius Δ_k

$\rightarrow$ If more than two levels are available $(r > 1)$, do this recursively

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



Notation:

- i: level index $(0 \leq i \leq r)$
- k: index of the current iteration within level i

Construction of the coarse local models

If $f_i \neq 0$ for $i = 0, \ldots, r-1$

- Impose first-order coherence via a correction term:

$$\boxed{g_{\text{low}} = R g_{\text{up}}}$$

- Impose second-order coherence$^{(*)}$ via two correction terms:

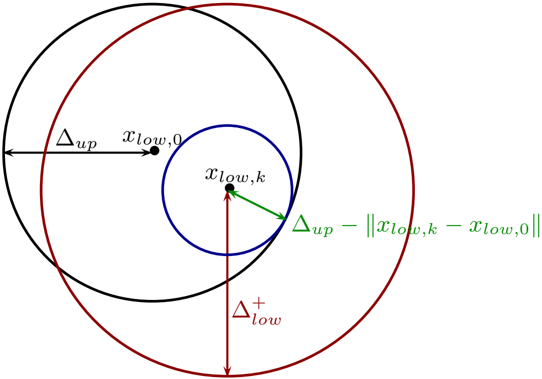
$$\boxed{g_{\text{low}} = R g_{\text{up}} \quad \text{and} \quad H_{\text{low}} = R H_{\text{up}} P}$$

$^{(*)}$ Not needed to derive first-order global convergence

If $f_i = 0$ for $i = 0, \ldots, r-1$

- Galerkin model: Restricted version of the quadratic model at the upper level

Preserving the trust-region constraint



$$\rightarrow \min\left[\Delta^{+}_{low},\ \Delta_{up} - \|x_{low,k} - x_{low,0}\|\right]$$

Note: Motivation to switch to ∞-norm

[Gratton, Mouffe, Toint, Weber Mendonça, to appear]

Use the coarse model whenever suitable

- When $\boxed{\|g_{\text{low}}\| \stackrel{\text{def}}{=} \|R g_{\text{up}}\| \geq \kappa \, \|g_{\text{up}}\|}$ (“Coarsening condition”)

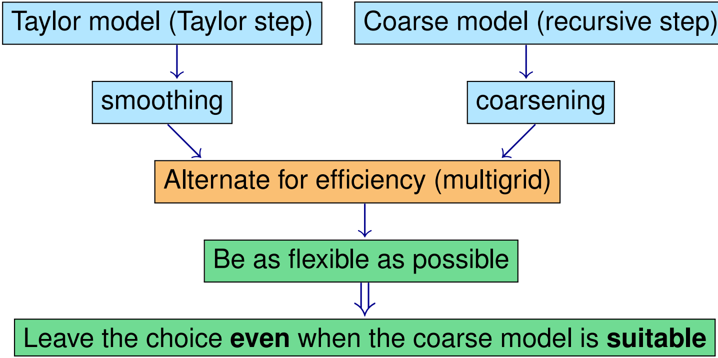
and

- When $\boxed{\|g_{\text{low}}\| \stackrel{\text{def}}{=} \|R g_{\text{up}}\| > \epsilon_{low}}$

and

- When $\boxed{i > 0}$

Use the coarse model whenever desirable



Recursive multilevel trust-region algorithm (RMTR)

At iteration k (until convergence):

- Choose either a Taylor or (if suitable) a coarse local model (first-order coherent):
  - Taylor model: compute a Taylor step (sufficient decrease condition OK)
  - Coarse local model: apply the algorithm recursively

    (sufficient decrease condition KO)
- Evaluate the change in the objective function
- If achieved decrease $\approx$ predicted decrease, then
  - accept the trial point
  - possibly enlarge the trust region

  else
  - keep the current point
  - shrink the trust region
- Impose current trust region $\subseteq$ upper level trust region

Global convergence

Based on the trust-region technology

- Uses the sufficient decrease argument (imposed in Taylor's iterations)
- Plus the coarsening condition $(\|Rg_{\text{up}}\| \geq \kappa \, \|g_{\text{up}}\|)$

Main result

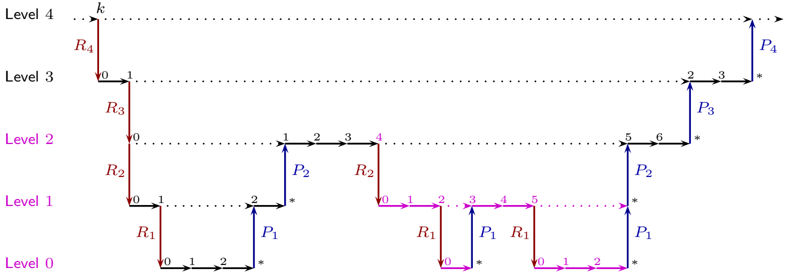
$$\boxed{\lim_{k \to \infty} \|g_{r,k}\| = 0}$$

[Gratton, Sartenaer, Toint, to appear]

Intermediate results

At iteration (i,k) we associate the set:

$$\mathcal{R}(i,k) \stackrel{\text{def}}{=} \{(j,\ell) \mid \text{iteration } (j,\ell) \text{ occurs within iteration } (i,k)\}$$



Let

$$\mathcal{V}(i,k) \stackrel{\text{def}}{=} \{ (j,\ell) \in \mathcal{R}(i,k) \mid \underbrace{\Delta m_{j,\ell} \geq \kappa \|g_{i,k}\| \Delta_{j,\ell}}_{\text{“sufficient decrease”}} \}$$

Then, at a non critical point and if the trust region is small enough:

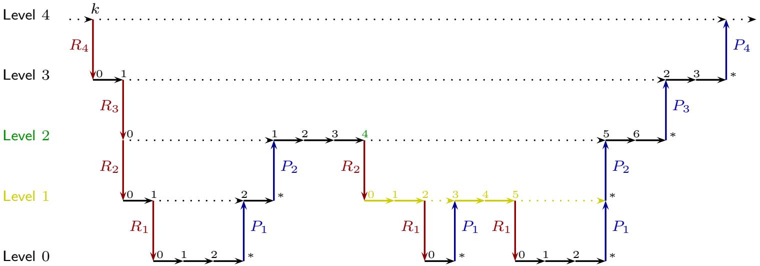
$$\mathcal{V}(i,k) = \mathcal{R}(i,k)$$

$\longrightarrow$ Back to “classical” trust-region arguments

Premature termination

For a recursive iteration (i, k):

A minimization sequence at level $i-1$ initiated at iteration (i, k)
denotes all successive iterations at level $i-1$
until a return is made to level i



Properties of RMTR

- Each minimization sequence contains at least one successful iteration
- Premature termination in that case does not affect the convergence results at the upper level

Which allows

- Stop a minimization sequence after a preset number of successful iterations
- Use fixed lower-iterations patterns like the V or W cycles in multigrid methods

A practical RMTR algorithm: Taylor iterations

At the coarsest level

- **Solve** using the exact Moré-Sorensen method (small dimension)

At finer levels

- **Smooth** using a smoothing technique from multigrid (to reduce the high frequency residual/gradient components)

SCM Smoothing

Adaptation of the Gauss-Seidel smoothing technique to optimization:

- Sequential Coordinate Minimization (SCM smoothing)

> Successive one-dimensional minimizations of the model along the coordinate axes when positive curvature

- Cost: 1 SCM smoothing cycle $\approx$ 1 matrix-vector product

Three issues

- How to $\boxed{\text{impose sufficient decrease in the model}}$?
- How to $\boxed{\text{impose the trust-region constraint}}$?
- What to $\boxed{\text{do if a negative curvature is encountered}}$?

Start the first SCM smoothing cycle

- by minimizing along the largest gradient component (enough to ensure sufficient decrease)

Perform (at most) p SCM smoothing cycles

- while inside the trust region (reasonable cost)

Terminate

- when an approximate minimizer is found (Stop)
- when the trust-region boundary is passed (Stop at the boundary)
- when a direction of negative curvature is encountered (move to the boundary and Stop)

Convergence to weak minimizers

SCM smoothing limits its exploration of the model's curvature to the coordinate axes $\rightarrow$ only guarantees asymptotic positive curvature:

- along the coordinate axes at the finest level $(i = r)$
- along the the prolongation of the coordinate axes at levels $i = 1, \ldots, r-1$
- along the prolongation of the coarsest subspace $(i = 0)$

"Weak" minimizers

Some numerical flavor

[Gratton, Mouffe, Sartenaer, Toint, Tomanos, in preparation]

All on Finest (**AF**)

Standard Newton trust-region algorithm (TCG)
Applied at the finest level

Multilevel on Finest (**MF**)

Algorithm RMTR
Applied at the finest level

Mesh Refinement (**MR**)

Standard Newton trust-region algorithm (TCG)
Applied successively from coarsest to finest level$^{(*)}$

Full Multilevel (**FM**)

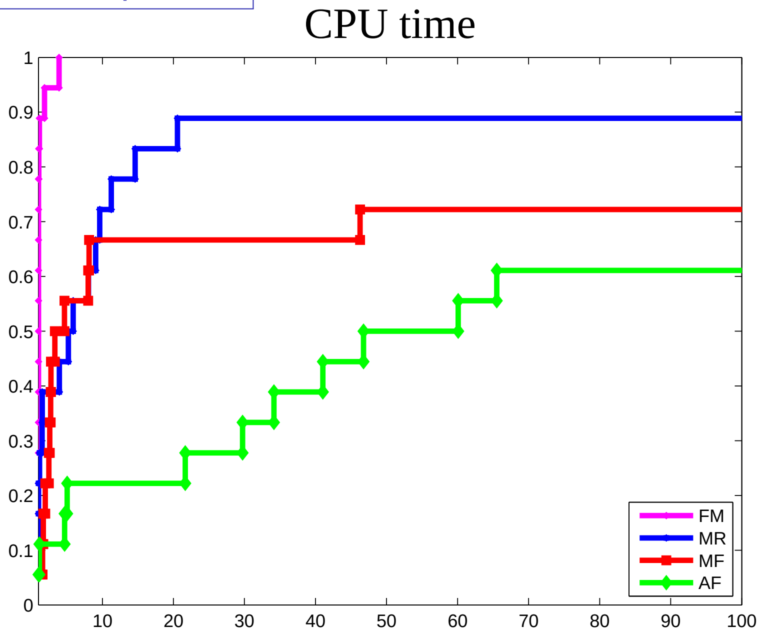
Algorithm RMTR
Applied successively from coarsest to finest level$^{(*)}$

$^{(*)}$ Starting point at level $i+1$ obtained by prolongating the solution at level i

Test problem characteristics

Problem name	n_r	r	Type	Bounds	Description
P2D	1.046.529	9	2-D, quadratic		Poisson model problem
P3D	250.047	5	3-D, quadratic		Poisson model problem
DEPT	1.046.529	9	2-D, quadratic		Elastic-plastic torsion problem
DPJB	1.046.529	9	2-D, quadratic	x	Journal bearing problem
DODC	1.046.529	9	2-D, convex		Optimal design problem
MINS-SB	1.046.529	9	2-D, convex		Minimium surface problem
MINS-OB	1.046.529	9	2-D, convex		Minimium surface problem
MINS-DMSA	1.046.529	9	2-D, convex		Minimium surface problem
IGNISC	3.969	5	2-D, convex		Combustion problem
DSSC	1.046.529	9	2-D, convex		Combustion problem
BRATU	1.046.529	9	2-D, convex		Combustion problem
MINS-BC	1.046.529	9	2-D, convex	x	Minimium surface problem
MEMBR	16.383	6	2-D, convex	x	Membrane problem
NCCS	7.938	6	2-D, nonconvex		Optimal control problem
NCCO	7.938	6	2-D, nonconvex		Optimal control problem
MOREBV	1.046.529	9	2-D, nonconvex		Boundary value problem

Performance profiles



CPU times

Problem name	AF	MF	MR	FM
P2D	1061.2	70.3	532.1	25.9
P3D	626.1	60.9	18.3	71.7
DEPT	1350.5	70.1	97.1	8.7
DPJB	3600.0	506.6	249.8	63.2
DODC	868.1	57.4	171.6	29.2
MINS-SB	3600.0	3600.0	3600.0	153.6
MINS-OB	1433.6	54.0	114.0	21.9
MINS-DMSA	1155.7	89.8	281.0	19.2
IGNISC	8.5	4.7	2.0	1.7
DSSC	3183.8	3600.0	116.1	12.1
BRATU	2020.7	1227.3	80.1	9.9
MINS-BC	2706.4	97.0	524.6	57.9
MEMBR	18.2	10.0	5.9	3.9
NCCS	146.1	2212.6	6.7	7.0
NCCO	145.6	3600.0	0.0	0.0
MOREBV	3600.0	1572.7	3600.0	34.0

Best Second best

In summary

- Successful merging of robustness and efficiency
- Still a lot to investigate
- Lots of applications