Multilevel Optimization

Annick Sartenaer

University of Namur, Belgium

SIAM Conference on Optimization, Boston, May 2008
The problem

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

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

\[ \rightarrow \text{Iterative search of a first-order critical point } x_* \text{ (s.t. } \nabla f(x_*) = 0) \]
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 \( x_* \)

\[ \rightarrow \text{Requires a } \text{globalization technique} \text{ in order to:} \]
  - 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
Assume now that a hierarchy of problem descriptions is available, linked by known operators.
Grid transfer operators

Restriction

\[ R_i : \mathbb{R}^{n_i} \rightarrow \mathbb{R}^{n_{i-1}} \]

Prolongation

\[ P_i : \mathbb{R}^{n_{i-1}} \rightarrow \mathbb{R}^{n_i} \]

\[ R_i = \sigma P_i^T \]
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

\[
\min_v \int_0^1 \int_0^1 \left( 1 + (\partial_x v)^2 + (\partial_y v)^2 \right)^{\frac{1}{2}} \, dx \, dy
\]

with the boundary conditions:

\[
\begin{cases}
  f(x), & y = 0, \quad 0 \leq x \leq 1 \\
  0, & x = 0, \quad 0 \leq y \leq 1 \\
  f(x), & y = 1, \quad 0 \leq x \leq 1 \\
  0, & x = 1, \quad 0 \leq y \leq 1
\end{cases}
\]

where

\[
f(x) = x \cdot (1 - x)
\]

\[\rightarrow\] 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

Efficiency – Robustness

Illustration within a trust-region framework

(Unconstrained case)
Past and recent developments

**Line-search**

- 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
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 ($\Delta f \approx \text{predicted decrease}$ ($\Delta m_k$)), then
  - accept the trial point ($x_{k+1} = x_k + s_k$)
  - possibly enlarge the trust region ($\Delta_k \uparrow$)

else

  - keep the current point ($x_{k+1} = x_k$)
  - shrink the trust region ($\Delta_k \downarrow$)
minimize:  \[ f(\alpha, \beta) = -10\alpha^2 + 10\beta^2 + 4\sin(\alpha\beta) - 2\alpha + \alpha^4 \]

Two local minima:  \((-2.20, 0.32) \text{ and } (2.30, -0.34)\)
\[ x_0 = (0.71, -3.27) \quad \text{and} \quad f(x_0) = 97.630 \]
<table>
<thead>
<tr>
<th>$k$</th>
<th>$\Delta_k$</th>
<th>$s_k$</th>
<th>$f(x_k + s_k)$</th>
<th>$\Delta f/\Delta m_k$</th>
<th>$x_{k+1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>(0.05, 0.93)</td>
<td>43.742</td>
<td>0.998</td>
<td>$x_0 + s_0$</td>
</tr>
<tr>
<td>$k$</td>
<td>$\Delta_k$</td>
<td>$s_k$</td>
<td>$f(x_k + s_k)$</td>
<td>$\Delta f / \Delta m_k$</td>
<td>$x_{k+1}$</td>
</tr>
<tr>
<td>-----</td>
<td>-----------</td>
<td>-----------</td>
<td>----------------</td>
<td>--------------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>(0.05, 0.93)</td>
<td>43.742</td>
<td>0.998</td>
<td>$x_0 + s_0$</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>(-0.62, 1.78)</td>
<td>2.306</td>
<td>1.354</td>
<td>$x_1 + s_1$</td>
</tr>
<tr>
<td>$k$</td>
<td>$\Delta_k$</td>
<td>$s_k$</td>
<td>$f(x_k + s_k)$</td>
<td>$\Delta f / \Delta m_k$</td>
<td>$x_{k+1}$</td>
</tr>
<tr>
<td>-----</td>
<td>------------</td>
<td>--------------</td>
<td>----------------</td>
<td>--------------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>(0.05, 0.93)</td>
<td>43.742</td>
<td>0.998</td>
<td>$x_0 + s_0$</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>(−0.62, 1.78)</td>
<td>2.306</td>
<td>1.354</td>
<td>$x_1 + s_1$</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>(3.21, 0.00)</td>
<td>6.295</td>
<td>−0.004</td>
<td>$x_2$</td>
</tr>
<tr>
<td>$k$</td>
<td>$\Delta_k$</td>
<td>$s_k$</td>
<td>$f(x_k + s_k)$</td>
<td>$\Delta f / \Delta m_k$</td>
<td>$x_{k+1}$</td>
</tr>
<tr>
<td>-----</td>
<td>------------</td>
<td>---------------</td>
<td>----------------</td>
<td>--------------------------</td>
<td>--------------------</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>(0.05, 0.93)</td>
<td>43.742</td>
<td>0.998</td>
<td>$x_0 + s_0$</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>(−0.62, 1.78)</td>
<td>2.306</td>
<td>1.354</td>
<td>$x_1 + s_1$</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>(3.21, 0.00)</td>
<td>6.295</td>
<td>−0.004</td>
<td>$x_2$</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>(1.90, 0.08)</td>
<td>−29.392</td>
<td>0.649</td>
<td>$x_2 + s_2$</td>
</tr>
</tbody>
</table>

![Contour plot](image1.png)

![Contour plot with circle](image2.png)
<table>
<thead>
<tr>
<th>$k$</th>
<th>$\Delta_k$</th>
<th>$s_k$</th>
<th>$f(x_k + s_k)$</th>
<th>$\Delta f / \Delta m_k$</th>
<th>$x_{k+1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>(0.05, 0.93)</td>
<td>43.742</td>
<td>0.998</td>
<td>$x_0 + s_0$</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>(−0.62, 1.78)</td>
<td>2.306</td>
<td>1.354</td>
<td>$x_1 + s_1$</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>(3.21, 0.00)</td>
<td>6.295</td>
<td>−0.004</td>
<td>$x_2$</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>(1.90, 0.08)</td>
<td>−29.392</td>
<td>0.649</td>
<td>$x_2 + s_2$</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>(0.32, 0.15)</td>
<td>−31.131</td>
<td>0.857</td>
<td>$x_3 + s_3$</td>
</tr>
</tbody>
</table>

The diagrams show contour plots of a function, with arrows indicating the direction of increase for each step $k$. The arrows point from $x_k$ to $x_{k+1}$, showing the change in $x$ after applying the correction $s_k$. The color and gradient indicate the rate of change, with darker colors representing higher values or steeper gradients.
<table>
<thead>
<tr>
<th>$k$</th>
<th>$\Delta_k$</th>
<th>$s_k$</th>
<th>$f(x_k + s_k)$</th>
<th>$\Delta f / \Delta m_k$</th>
<th>$x_{k+1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>(0.05, 0.93)</td>
<td>43.742</td>
<td>0.998</td>
<td>$x_0 + s_0$</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>(−0.62, 1.78)</td>
<td>2.306</td>
<td>1.354</td>
<td>$x_1 + s_1$</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>(3.21, 0.00)</td>
<td>6.295</td>
<td>−0.004</td>
<td>$x_2$</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>(1.90, 0.08)</td>
<td>−29.392</td>
<td>0.649</td>
<td>$x_2 + s_2$</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>(0.32, 0.15)</td>
<td>−31.131</td>
<td>0.857</td>
<td>$x_3 + s_3$</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>(−0.03, −0.02)</td>
<td>−31.176</td>
<td>1.009</td>
<td>$x_4 + s_4$</td>
</tr>
<tr>
<td>$k$</td>
<td>$\Delta_k$</td>
<td>$s_k$</td>
<td>$f(x_k + s_k)$</td>
<td>$\Delta f / \Delta m_k$</td>
<td>$x_{k+1}$</td>
</tr>
<tr>
<td>-----</td>
<td>-----------</td>
<td>--------------</td>
<td>----------------</td>
<td>--------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>(0.05, 0.93)</td>
<td>43.742</td>
<td>0.998</td>
<td>$x_0 + s_0$</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>(-0.62, 1.78)</td>
<td>2.306</td>
<td>1.354</td>
<td>$x_1 + s_1$</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>(3.21, 0.00)</td>
<td>6.295</td>
<td>-0.004</td>
<td>$x_2$</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>(1.90, 0.08)</td>
<td>-29.392</td>
<td>0.649</td>
<td>$x_2 + s_2$</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>(0.32, 0.15)</td>
<td>-31.131</td>
<td>0.857</td>
<td>$x_3 + s_3$</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>(-0.03, -0.02)</td>
<td>-31.176</td>
<td>1.009</td>
<td>$x_4 + s_4$</td>
</tr>
<tr>
<td>6</td>
<td>8</td>
<td>(-0.02, 0.00)</td>
<td>-31.179</td>
<td>1.013</td>
<td>$x_5 + s_5$</td>
</tr>
</tbody>
</table>
Path of iterates:

From another $x_0$: 
What makes it work?

\[
\begin{align*}
\begin{cases}
\text{(approx.) minimize} & m_k(x_k + s) \\
\text{subject to} & \|s\| \leq \Delta_k
\end{cases}
\end{align*}
\]

\[
\begin{align*}
\left( g_k = \nabla m_k(x_k) = \nabla f(x_k) \right)
\end{align*}
\]

\[
\begin{align*}
\begin{cases}
\text{minimize} & t \geq 0 \\
\text{subject to} & t \|g_k\| \leq \Delta_k
\end{cases}
\end{align*}
\]

Model decrease at the Cauchy point \( x_k^C \)

(Best decrease of the model within the trust region along the steepest descent direction \(-g_k\)
\[ m_k(x_k) - m_k(x_k^c) \geq \frac{1}{2} \|g_k\| \min \left[ \frac{\|g_k\|}{\beta_k}, \Delta_k \right] \]

Illustration from [Conn, Gould, Toint, 2000]

positive curvature and minimum inside

positive curvature and minimum outside

negative curvature

\[ \beta_k = \text{upper bound on the curvature of } m_k \]
First-order convergence

Sufficient decrease condition:

$$m_k(x_k) - m_k(x_k + s_k) \geq \kappa (m_k(x_k) - m_k(x_k^C))$$

$$\Rightarrow \lim_{k \to \infty} \| \nabla f(x_k) \| = 0$$
What makes it fast?

\[ m_k(x_k + s) = f(x_k) + g_k^T s + \frac{1}{2} s^T H_k s \]

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

Any **global minimizer** \( s^* \) of

\[
\begin{align*}
\text{minimize}_{s \in \mathbb{R}^n} & \quad m_k(x_k + s) \\
\text{subject to} & \quad \|s\|_2 \leq \Delta_k
\end{align*}
\]

satisfies:

\[
(H_k + \lambda^* I) s^* = -g_k
\]

where \( H_k + \lambda^* I \) is pos. (semi)def., \( \lambda^* \geq 0 \) and \( \lambda^*(\|s^*\|_2 - \Delta_k) = 0 \)
For $\Delta_k$ fixed, find $\lambda \geq \max\{0, -\lambda_{\text{min}}(H_k)\}$ such that:

- $H_k + \lambda I$ is positive semidefinite

- $s(\lambda) = -(H_k + \lambda I)^{-1} g_k$ 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:

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

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

$$\begin{align*}
\begin{cases}
 \text{(approx) min}_{s \in \mathbb{R}^n} & m_k(x_k + s) \\
 \text{subject to} & \|s\|_2 \leq \Delta_k
\end{cases}
\end{align*}$$
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)
Trust-region methods [Conn, Gould, Toint, 2000]
On the side of multigrid methods

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

\[ Ax = b \quad \xrightarrow{\sim} \quad Ae = r \]

(residual equation)

where

- \[ e = x_\ast - \tilde{x} \] (error)
- \[ x_\ast \] (exact solution)
- \[ r = b - A\tilde{x} \] (residual)
- \[ \tilde{x} \] (approximation)

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

![Fourier modes graph](image_url)
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**

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

→ Solve the equations of the linear system one by one
Very effective methods at “smoothing”, i.e., eliminating the high-frequency (oscillatory) components of the error:

But they leave the low-frequency (smooth) components relatively unchanged
Assume now (two levels):

- **A fine grid \((f)\) description**
  \[ Ae = r \quad \rightarrow \quad A^f e^f = r^f \]

- **A coarse grid \((c)\) description**
  \[ A^c e^c = r^c \]

- **Linked by transfer operators**
  \[ A^c = RA^f P, \quad e^c = Re^f, \quad r^c = Rr^f \]
Coarse grid principle

Smooth error modes on a fine grid “look less smooth” on a coarse grid

→ 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

Fine $e$ → Smooth fine $e$ ↓ $R$ → Oscil. coarse $e$ → Smooth coarse $e$ → Smaller oscil. fine $e$ → Smaller smooth fine $e$
Smoothing on fine grid only:

Two-grid correction scheme:

\[ k = 0 \quad k = 10 \quad k = 100 \]
Recursive use to annihilate oscillatory error level by level \( (O(n)) \)
V-cycle

\[ k \quad \cdots \cdots \quad k + 1 \]

\[ 0 \quad 1 \quad 2 \quad * \]

Smoothing
W-cycle

Smoothing
Mesh Refinement

- Solve the problem on the coarsest level

  ⇒ Good starting point for the next fine level

- Do the same on each level

  ⇒ 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

Efficiency – Robustness

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 $f$:

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

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

Terminology: a particular $i$ is referred to as a level
The idea

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

\[\downarrow \text{ or (whenever suitable and desirable)}\]

\[\text{at } x_k:\text{ compute } \nabla f_r(x_k) \text{ (possibly } H_k) \quad \text{trial step } s_k\]

\[\text{Restriction } \downarrow R \quad P \uparrow \text{Prolongation}\]

\[\text{use } f_{r-1} \text{ to construct a } \textbf{coarse local} \text{ model of } f_r \text{ and minimize it within the trust region of radius } \Delta_k\]

\[\rightarrow \text{ If more than two levels are available } (r > 1), \text{ do this recursively}\]
Example of recursion with 5 levels \((r = 4)\)

Notation:
\[
\begin{align*}
  i & : \text{level index} \quad (0 \leq i \leq r) \\
  k & : \text{index of the current iteration within level } i
\end{align*}
\]
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:
  \[
  g_{\text{low}} = R g_{\text{up}}
  \]

- Impose second-order coherence\(^(*)\) via two correction terms:
  \[
  g_{\text{low}} = R g_{\text{up}} \quad \text{and} \quad H_{\text{low}} = R H_{\text{up}} P
  \]

\(^(*)\) 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

\[ \Delta_{up} x_{low,0} \]

\[ \Delta^+_{low} \]

\[ \Delta_{up} - \|x_{low,k} - x_{low,0}\| \]

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

Note: Motivation to switch to \( \infty \)-norm

[Gratton, Mouffe, Toint, Weber Mendonça, to appear]
Use the coarse model whenever suitable

► When \( \| g_{\text{low}} \| \overset{\text{def}}{=} \| Rg_{\text{up}} \| \geq \kappa \| g_{\text{up}} \| \) ("Coarsening condition")

and

► When \( \| g_{\text{low}} \| \overset{\text{def}}{=} \| Rg_{\text{up}} \| > \epsilon_{\text{low}} \)

and

► When \( i > 0 \)
Use the coarse model whenever desirable

Taylor model (Taylor step)  Coarse model (recursive step)

↓  ↓

smoothing  coarsening

Alternate for efficiency (multigrid)

Be as flexible as possible

Leave the choice even when the coarse model is suitable
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**  \( \| R_{g_{up}} \| \geq \kappa \| g_{up} \| \)

Main result

\[
\lim_{k \to \infty} \| g_{r,k} \| = 0
\]

[Gratton, Sartenaer, Toint, to appear]
Intermediate results

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

\[
\mathcal{R}(i, k) \overset{\text{def}}{=} \{(j, \ell) \mid \text{iteration } (j, \ell) \text{ occurs within iteration } (i, k)\}
\]
Let

\[ V(i, k) \overset{\text{def}}{=} \{ (j, \ell) \in R(i, k) \mid \frac{\Delta m_{j,\ell}}{\Delta} \geq \kappa \|g_{i,k}\| \Delta_{j,\ell} \} \]

“sufficient decrease”

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

\[ V(i, k) = R(i, k) \]

\[ \rightarrow \text{ Back to “classical” trust-region arguments} \]
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)
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 impose sufficient decrease in the model?

- How to impose the trust-region constraint?

- What to 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)
SCM smoothing limits its exploration of the model’s curvature to the coordinate axes → only guarantees asymptotic positive curvature:

- along the coordinate axes at the finest level \((i = r)\)
- along 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

<table>
<thead>
<tr>
<th>Problem name</th>
<th>( n_r )</th>
<th>( r )</th>
<th>Type</th>
<th>Bounds</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>P2D</td>
<td>1.046.529</td>
<td>9</td>
<td>2-D, quadratic</td>
<td></td>
<td>Poisson model problem</td>
</tr>
<tr>
<td>P3D</td>
<td>250.047</td>
<td>5</td>
<td>3-D, quadratic</td>
<td></td>
<td>Poisson model problem</td>
</tr>
<tr>
<td>DEPT</td>
<td>1.046.529</td>
<td>9</td>
<td>2-D, quadratic</td>
<td></td>
<td>Elastic-plastic torsion problem</td>
</tr>
<tr>
<td>DPJB</td>
<td>1.046.529</td>
<td>9</td>
<td>2-D, quadratic</td>
<td>x</td>
<td>Journal bearing problem</td>
</tr>
<tr>
<td>DODC</td>
<td>1.046.529</td>
<td>9</td>
<td>2-D, convex</td>
<td></td>
<td>Optimal design problem</td>
</tr>
<tr>
<td>MINS-SB</td>
<td>1.046.529</td>
<td>9</td>
<td>2-D, convex</td>
<td></td>
<td>Minimum surface problem</td>
</tr>
<tr>
<td>MINS-OB</td>
<td>1.046.529</td>
<td>9</td>
<td>2-D, convex</td>
<td></td>
<td>Minimum surface problem</td>
</tr>
<tr>
<td>MINS-DMSA</td>
<td>1.046.529</td>
<td>9</td>
<td>2-D, convex</td>
<td></td>
<td>Minimum surface problem</td>
</tr>
<tr>
<td>IGNISC</td>
<td>3.969</td>
<td>5</td>
<td>2-D, convex</td>
<td></td>
<td>Combustion problem</td>
</tr>
<tr>
<td>DSSC</td>
<td>1.046.529</td>
<td>9</td>
<td>2-D, convex</td>
<td></td>
<td>Combustion problem</td>
</tr>
<tr>
<td>BRATU</td>
<td>1.046.529</td>
<td>9</td>
<td>2-D, convex</td>
<td></td>
<td>Combustion problem</td>
</tr>
<tr>
<td>MINS-BC</td>
<td>1.046.529</td>
<td>9</td>
<td>2-D, convex</td>
<td>x</td>
<td>Minimum surface problem</td>
</tr>
<tr>
<td>MEMBR</td>
<td>16.383</td>
<td>6</td>
<td>2-D, convex</td>
<td>x</td>
<td>Membrane problem</td>
</tr>
<tr>
<td>NCCS</td>
<td>7.938</td>
<td>6</td>
<td>2-D, nonconvex</td>
<td></td>
<td>Optimal control problem</td>
</tr>
<tr>
<td>NCCO</td>
<td>7.938</td>
<td>6</td>
<td>2-D, nonconvex</td>
<td></td>
<td>Optimal control problem</td>
</tr>
<tr>
<td>MOREBV</td>
<td>1.046.529</td>
<td>9</td>
<td>2-D, nonconvex</td>
<td></td>
<td>Boundary value problem</td>
</tr>
</tbody>
</table>
## CPU times

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

**Best**

**Second best**
In summary

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