# A Retrospective Trust-Region Method for Unconstrained Optimization

F. Bastin[1] V. Malmedy[2,3] M. Mouffe[4] Ph. Toint[2]
D. Tomanos[2,5]

[1]Computing Science and Operational Research Department, University of Montreal
[2]Department of Mathematics, FUNDP [3]F.R.S.-FNRS, Research Fellow
[4]CERFACS, Toulouse [5]FRIA, Research Fellow
(philippe.toint@fundp.ac.be)

AMA Workshop, Hong Kong, June 2008

# The problem

Unconstrained optimization:

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

with objective function $f : \mathbf{R}^n \to \mathbf{R}$

- nonlinear, twice-continuously differentiable, and bounded below
- no convexity assumption

# Basic Trust-Region method (BTR)

Until convergence:

1. choose a local model $m_k$ of the objective $f$ around $x_k$
2. compute a trial point $x_k + s_k$ that decreases the model $m_k$ within a trust-region $\|s_k\| \leq \Delta_k$
3. compute the reduction ratio

$$\rho_k := \frac{f(x_k) - f(x_k + s_k)}{m_k(x_k) - m_k(x_k + s_k)}$$

4. if $m_k$ and $f$ agree at $x_k + s_k$, *i.e.* $\rho_k \geq \eta_1$

   then

   - accept trial point: $x_{k+1} = x_k + s_k$
   - update the trust region radius: $\Delta_{k+1} \in \begin{cases} [\Delta_k, \infty) & \text{if } \rho_k \geq \eta_2 \\ [\gamma_2 \Delta_k, \Delta_k) & \text{if } \rho_k \in [\eta_1, \eta_2) \end{cases}$

   else

   - reject trial point: $x_{k+1} = x_k$
   - reduce the trust region radius: $\Delta_{k+1} \in [\gamma_1 \Delta_k, \gamma_2 \Delta_k)$

with $0 < \eta_1 \leq \eta_2 < 1$ and $0 < \gamma_1 \leq \gamma_2 < 1$

# Roles of reduction ratio — Main idea of new method

In BTR, the reduction ratio $\rho_k$ plays two roles:

1. acceptance of the trial point $x_k + s_k$
2. control of the trust-region radius update

Idea: distinguish these two roles, since:

1. acceptance step based on how well the current model $m_k$ predicts the decrease of the function $f$ at $x_k + s_k$
2. updated radius used to define where the new model $m_{k+1}$ is trusted to agree with the function $f$ around $x_k + s_k$

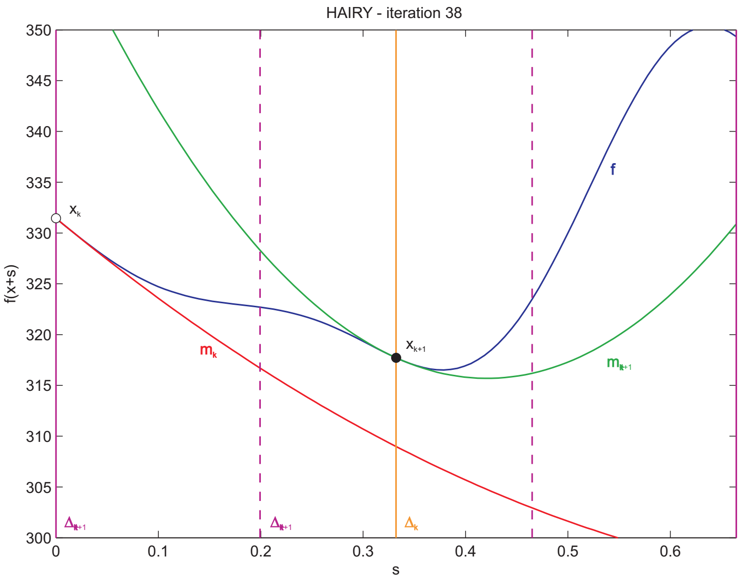
# Retrospective Trust-Region method (RTR)

Until convergence:

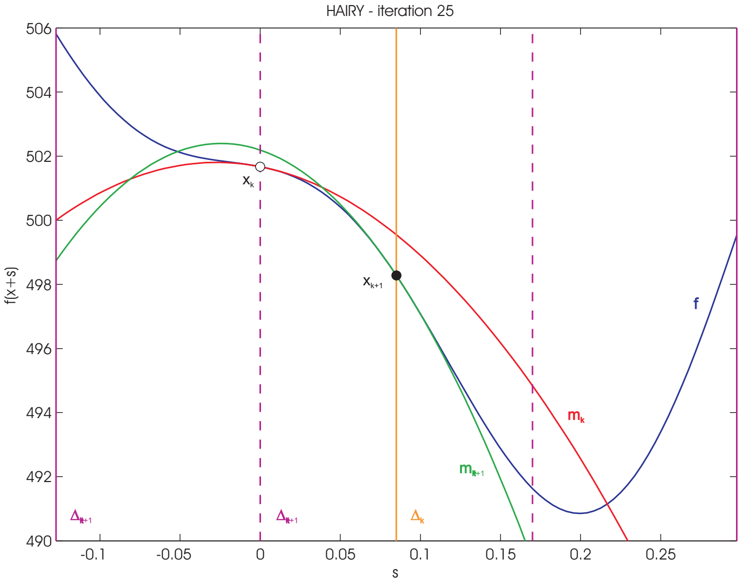
1. choose a local model $m_k$ of the objective $f$ around $x_k$
2. if former trial point was rejected
   then reduce the trust-region radius: $\Delta_k \in [\gamma_1 \Delta_{k-1}, \gamma_2 \Delta_{k-1})$
   else compute
   $$\tilde{\rho}_k := \frac{f(x_{k-1}) - f(x_k)}{m_k(x_{k-1}) - m_k(x_k)}$$
   and update the trust-region radius:
   $$\Delta_k \in \begin{cases} [\Delta_{k-1}, \infty) & \text{if } \tilde{\rho}_k \geq \tilde{\eta}_2 \\ [\gamma_2 \Delta_{k-1}, \Delta_{k-1}) & \text{if } \tilde{\rho}_k \in [\tilde{\eta}_1, \tilde{\eta}_2) \end{cases}$$
3. compute a trial point $x_k + s_k$ decreasing the model $m_k$ within $\|s_k\| \leq \Delta_k$
4. compute the reduction ratio $\rho_k := \frac{f(x_k) - f(x_k + s_k)}{m_k(x_k) - m_k(x_k + s_k)}$
5. if $\rho_k \geq \eta_1$, accept trial point: $x_{k+1} = x_k + s_k$;
   otherwise reject trial point: $x_{k+1} = x_k$

with $0 < \eta_1 < 1$, $0 < \tilde{\eta}_1 \leq \tilde{\eta}_2 < 1$ and $0 < \gamma_1 \leq \gamma_2 < 1$

# Graphically. . . (1)



# Graphically. . . (2)



# Convergence theory

RTR no more covered by classical theory
$\Rightarrow$ need of an adapted convergence theory

**Assume:**

- $\nabla_{xx} f$ and $\nabla_{xx} m_k$ uniformly bounded
- first-order coherent models: $\nabla_x f(x_k) = \nabla_x m_k(x_k)$
- sufficient decrease condition (at least a fraction of Cauchy point):

$$m_k(x_k) - m_k(x_k + s_k) \geq \gamma \|g_k\| \min(\|g_k\|/\beta_k, \Delta_k)$$

# First-order convergence

Where changes occurs?

Let $\delta_k m := m(x_k) - m(x_{k+1})$ be the reduction of model $m$ at iteration $k$. Then

$$|\delta_k m_k - \delta_k m_{k+1}| \leq \kappa \Delta_k^2.$$

If $g_k \neq 0$ and $\Delta_k \leq \zeta \|g_k\|$, then iteration $k$ is successful and $\Delta_k$ grows.

Finally, same results:

If only finitely many successful iterations,
then after some time, $x_k = x_*$ which is first-order critical.

$$\lim_{k \to \infty} \|\nabla_x f(x_k)\| = 0.$$

# Second-order convergence (1)

**Assume moreover:**

- asymptotically second-order coherent models near first-order critical points:

$$\|\nabla_{xx} f(x_k) - \nabla_{xx} m_k(x_k)\| \to 0 \quad \text{when} \quad \|g_k\| \to 0$$

Where changes occurs?

Suppose that $m_{k_i}(x_{k_i}) - m_{k_i}(x_{k_i} + s_{k_i}) \geq \nu \|s_{k_i}\|^2$ and that $s_{k_i} \to 0$.
Then iteration $k$ is successful and $\Delta_{k_i}$ grows.

# Second-order convergence (2)

**Assume furthermore:**

- $\nabla_{xx} m_k$ Lipschitz continuous
- if $\tau_k := \lambda_{\min}(\nabla_{xx} m_k) < 0$, then

$$m_k(x_k) - m_k(x_k + s_k) \geq \xi |\tau_k| \min(\tau_k^2, \Delta_k^2)$$

Finally, same results:

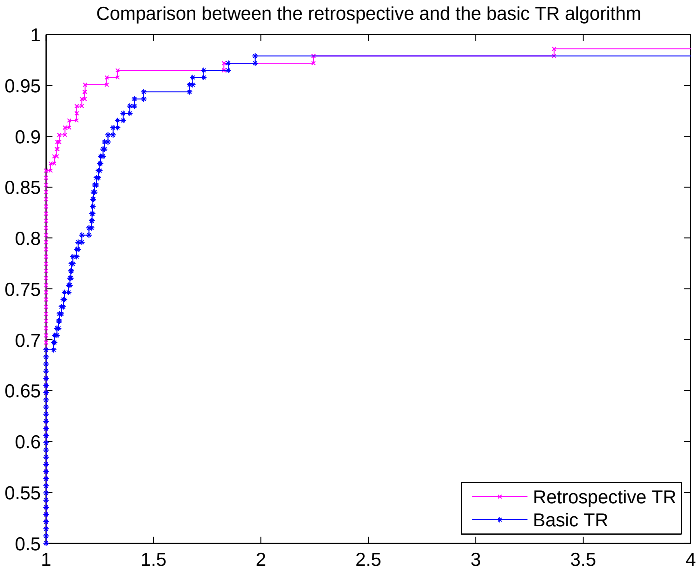
Suppose that $\{x_k\}$ remains in a compact set.
Then there exists at least one limit point $x_*$ that is second-order critical.

Suppose that $\Delta_{k+1} \in [\gamma_3 \Delta_k, \gamma_4 \Delta_k]$ whenever $\tilde{\rho}_k \geq \tilde{\eta}_2$ (with $\gamma_4 \geq \gamma_3 > 1$).
Then every limit point $x_*$ is second-order critical.

# Numerical experiments

- 146 unconstrained problems from CUTEr library (Gould, Orban, Toint, 2003) with size between 2 and 500
- matlab implementation
- classical parameters for TR as advised by Conn, Gould, Toint (2000)
- exact quadratic model
- subproblem solved with More-Sorensen method
- stopping criterion: $\|g_k\| \leq 10^{-5}$ or more than $10^5$ iterations

# Performance profile



# Conclusion and perspectives

## Conclusions

- exploitation of the most recent model information
- first- and second-order convergence theory
- improved numerical performances
- no supplementary cost

## Perspectives

- Stochastic programming (dynamic accuracy on the objective function computation)
- Combination with ACO methods?

Thank you for your attention!