Lecture 9: Trust-region methods for unconstrained optimization (continued)

Coralia Cartis, Mathematical Institute, University of Oxford

C6.2/B2: Continuous Optimization

The Cauchy point of the (TR) subproblem

• recall the steepest descent method has strong (theoretical) global convergence properties; same will hold for TR method with SD direction.

"minimal" condition of "sufficient decrease" in the model: require

$$m_k(s^k) \le m_k(s_c^k) \text{ and } \|s^k\| \le \Delta_k,$$

where $s_c^k := -\alpha_c^k \nabla f(x^k)$, with

$$\alpha_c^k := \arg\min_{\alpha>0} m_k(-\alpha\nabla f(x^k)) \text{ subject to } \|\alpha\nabla f(x^k)\| \le \Delta_k.$$

[i.e. a linesearch along steepest descent direction is applied to m_k at x^k and is restricted to the trust region.] Easy:

$$\alpha_c^k := \arg\min_{\alpha} m_k(-\alpha\nabla f(x^k)) \text{ subject to } 0 < \alpha \le \frac{\Delta_k}{\|\nabla f(x^k)\|}.$$

• $y_c^k := x^k + s_c^k$ is the Cauchy point.

Computation of the Cauchy point

Computation of the Cauchy point: find α_c^k global solution of

$$\min_{\alpha>0} m_k(-\alpha\nabla f(x^k)) \text{ subject to } \|\alpha\nabla f(x^k)\| \leq \Delta_k,$$

where $m_k(s) = f(x_k) + s^T\nabla f(x^k) + \frac{1}{2}s^T\nabla^2 f(x^k)s$, & $\nabla f(x^k) \neq 0$.

- $\|\alpha\nabla f(x^k)\| \leq \Delta_k$ & $\alpha > 0 \Leftrightarrow 0 < \alpha \leq \dfrac{\Delta_k}{\|\nabla f(x^k)\|} := \overline{\alpha}$.
- $\psi(\alpha) := m_k(-\alpha\nabla f(x^k)) = f(x^k) - \alpha\|\nabla f(x^k)\|^2 + \dfrac{\alpha^2}{2}h^k$,

where $h^k := \nabla f(x^k)^T\nabla^2 f(x^k)\nabla f(x^k)$.

Calculation of the Cauchy point

$\nabla f(x^k) \neq 0$: α_c^k solution of: $\min_{\alpha > 0} m_k(-\alpha \nabla f(x^k)) := \Psi(\alpha)$ subject to $\|\alpha \nabla f(x^k)\| \leq \Delta_k$.

$\| \ \alpha > 0, \ \nabla f(x^k) \neq 0$

$\Rightarrow$ $0 < \alpha \leq \dfrac{\Delta_k}{\|\nabla f(x^k)\|}$

$$\Psi(\alpha) = f(x^k) + \underbrace{(-\alpha \nabla f(x^k))}_{= s \text{ in } m_k(s)}{}^T \nabla f(x^k) + \frac{1}{2}(-\nabla f(x^k)\alpha)^T \nabla^2 f(x^k)(-\alpha \nabla f(x^k))$$

$$= f(x^k) - \alpha \underbrace{\nabla f(x^k)^T \nabla f(x^k)}_{= \|\nabla f(x^k)\|^2} + \frac{1}{2}\alpha^2 \underbrace{\nabla f(x^k)^T \nabla^2 f(x^k) \nabla f(x^k)}_{:= h^k}$$

To calculate α_c^k must solve:

$$\min_{\alpha \in \mathbb{R}} \Psi(\alpha) = f(x^k) - \alpha \|\nabla f(x^k)\|^2 + \frac{\alpha^2}{2} h^k \quad \underline{\underline{\text{s.t.}}} \quad \underbrace{0 < \alpha \leq \frac{\Delta_k}{\|\nabla f(x^k)\|} := \bar{\alpha}}_{\alpha \in (0, \bar{\alpha}]}$$

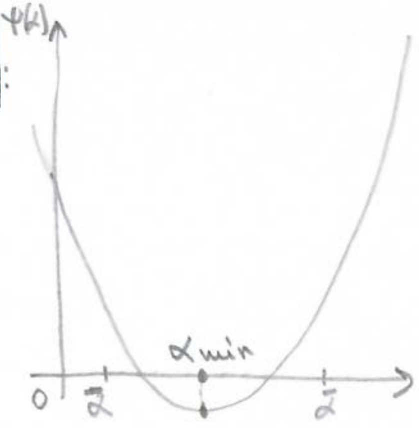
$\uparrow$ $\Psi(\alpha) = m_k(-\alpha \nabla f(x^k))$

quadratic function in α

$\Psi'(\alpha) = -\|\nabla f(x^k)\|^2 + \alpha h^k \Rightarrow \Psi'(0) < 0 \Rightarrow \Psi$ strictly decreasing at $\alpha = 0$.

$\Rightarrow$ two possible cases:

$h^k > 0$:



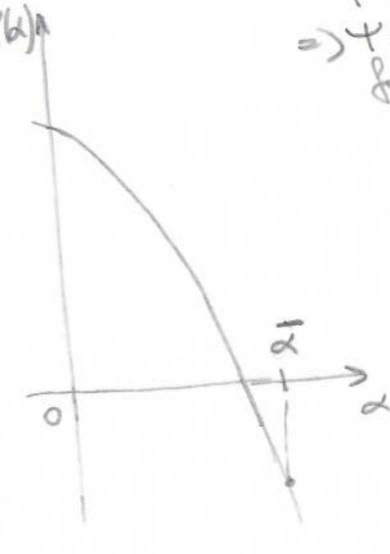
Ψ convex

$\Rightarrow$ value of α_c^k depends on that of $\bar{\alpha}$ and α_{min}.

$\Rightarrow$ $\boxed{\alpha_c^k = \min(\alpha_{min}, \bar{\alpha})}$

$\Psi'(\alpha_{min}) = 0 \Rightarrow \alpha_{min} = \dfrac{\|\nabla f(x^k)\|^2}{h^k}$

$h^k \leq 0$:



Ψ concave/linear

$\Rightarrow \Psi$ strictly decreasing for all $\alpha > 0$

$\Rightarrow \Psi(\alpha) \geq \Psi(\bar{\alpha}) \quad \forall \alpha \in (0, \bar{\alpha}]$

$\Rightarrow$ $\boxed{\alpha_c^k = \bar{\alpha}}$

Computation of the Cauchy point

Computation of the Cauchy point: find α_c^k global solution of

$$\min_{\alpha>0} m_k(-\alpha\nabla f(x^k)) \text{ subject to } \|\alpha\nabla f(x^k)\| \leq \Delta_k,$$

where $m_k(s) = f(x_k) + s^T\nabla f(x^k) + \frac{1}{2}s^T\nabla^2 f(x^k)s$, & $\nabla f(x^k) \neq 0$.

- $\|\alpha\nabla f(x^k)\| \leq \Delta_k$ & $\alpha > 0 \Leftrightarrow 0 < \alpha \leq \dfrac{\Delta_k}{\|\nabla f(x^k)\|} := \overline{\alpha}$.
- $\psi(\alpha) := m_k(-\alpha\nabla f(x^k)) = f(x^k) - \alpha\|\nabla f(x^k)\|^2 + \dfrac{\alpha^2}{2}h^k$,
  where $h^k := \nabla f(x^k)^T\nabla^2 f(x^k)\nabla f(x^k)$.
- $\psi'(0) = -\|\nabla f(x^k)\|^2 < 0$ so ψ decreasing from $\alpha = 0$ for suff. small α; thus $\alpha_c^k > 0$.
  - $h^k > 0$: $\alpha_{\min} := \dfrac{\|\nabla f(x^k)\|^2}{h^k} = \arg\min_{\alpha>0}\psi(\alpha)$.
    $\Longrightarrow \quad \alpha_c^k = \min(\alpha_{\min}, \overline{\alpha})$.
  - $h^k \leq 0$: $\psi(\alpha)$ unbounded below on $\mathbb{R}$ and so $\alpha_c^k = \overline{\alpha}$.

Proof of global convergence of the GTR method

Lemma 12: (Cauchy model decrease) In GTR with Cauchy decrease $m_k(s^k) \leq m_k(s_c^k)$ for all k, we have the model decrease for each k,

$$\begin{aligned} f(x^k) - m_k(s^k) &\geq f(x^k) - m_k(s_c^k) \\ &\geq \tfrac{1}{2}\|\nabla f(x^k)\| \min\left\{\Delta_k, \frac{\|\nabla f(x^k)\|}{\|\nabla^2 f(x^k)\|}\right\} \end{aligned}$$

Proof of Lemma 12. (Recall Computation of the Cauchy point)
If $h^k \leq 0$, then the definition of ψ implies
$m_k(s_c^k) = m_k(-\alpha_c^k \nabla f(x^k)) = \psi(\alpha_c^k) \leq f(x^k) - \alpha_c^k \|\nabla f(x^k)\|^2$.
In this case, we also have $\alpha_c^k = \overline{\alpha} = \dfrac{\Delta_k}{\|\nabla f(x^k)\|}$ and so

$$f(x^k) - m_k(s_c^k) \geq \alpha_c^k \|\nabla f(x^k)\|^2 = \tfrac{\Delta_k}{\|\nabla f(x^k)\|}\|\nabla f(x^k)\|^2 = \Delta_k \|\nabla f(x^k)\|.$$

Note that $\Delta_k \|\nabla f(x^k)\| \geq \frac{1}{2}\|\nabla f(x^k)\| \min\{\Delta_k, \ldots\}$.

Proof of global convergence of the GTR method

Proof of Lemma 12 (continued).

Else, $h^k > 0$; then $\alpha_c^k = \min\{\alpha_{\min}, \overline{\alpha}\}$ where $\alpha_{\min} = \|\nabla f(x^k)\|^2/h^k$. Assume first that $\alpha_c^k = \overline{\alpha}$. Then $\alpha_c^k \leq \alpha_{\min}$, which from the expression of $\alpha_{\min}$, implies $\alpha_c^k h^k \leq \|\nabla f(x^k)\|^2$. Now use this bound in the expression of $\psi(\alpha)$, namely,

$$\begin{aligned} f(x^k) - m_k(s_c^k) &= f(x^k) - \psi(\alpha_c^k) = \alpha_c^k \|\nabla f(x^k)\|^2 - \frac{(\alpha_c^k)^2}{2} h^k \\ &= \alpha_c^k \|\nabla f(x^k)\|^2 - \frac{\alpha_c^k}{2}(\alpha_c^k h^k) \\ &\geq (\alpha_c^k - \frac{\alpha_c^k}{2})\|\nabla f(x^k)\|^2 = \frac{\alpha_c^k}{2}\|\nabla f(x^k)\|^2, \end{aligned}$$

and using the expression of $\overline{\alpha}$, we deduce

$$f(x^k) - m_k(s_c^k) \geq \frac{\Delta_k}{2\|\nabla f(x^k)\|}\|\nabla f(x^k)\|^2 = \frac{1}{2}\Delta_k \|\nabla f(x^k)\|.$$

Note that $\frac{1}{2}\Delta_k \|\nabla f(x^k)\| \geq \frac{1}{2}\|\nabla f(x^k)\| \min\{\Delta_k, \ldots\}$.

Proof of global convergence of the GTR method

Proof of Lemma 12 (continued).

Finally, when $h^k > 0$, let $\alpha_c^k = \alpha_{\min} = \|\nabla f(x^k)\|^2/h^k$. Replacing this value in the model decrease we get

$$f(x^k) - m_k(s_c^k) = \alpha_c^k \|\nabla f(x^k)\|^2 - \frac{(\alpha_c^k)^2}{2} h^k = \frac{\|\nabla f(x^k)\|^4}{2h^k},$$

and further, by Cauchy-Schwarz and Rayleigh quotient inequalities (recall Pb 6, Sheet 2),

$$\begin{aligned} \frac{\|\nabla f(x^k)\|^4}{2h^k} &= \frac{\|\nabla f(x^k)\|^4}{2(\nabla f(x^k))^T \nabla^2 f(x^k) \nabla f(x^k)} \\ &\geq \frac{\|\nabla f(x^k)\|^4}{2\|\nabla^2 f(x^k)\| \cdot \|\nabla f(x^k)\|^2} \\ &\geq \frac{\|\nabla f(x^k)\|^2}{2\|\nabla^2 f(x^k)\|} \end{aligned}$$

Thus

$$f(x^k) - m_k(s_c^k) \geq \frac{\|\nabla f(x^k)\|^2}{2\|\nabla^2 f(x^k)\|} \geq \frac{1}{2}\|\nabla f(x^k)\| \min\{\ldots, \frac{\|\nabla f(x^k)\|}{\|\nabla^2 f(x^k)\|}\}.$$

□

A Generic Trust Region (GTR) method

Given $\Delta_0 > 0$, $x^0 \in \mathbb{R}^n$, $\epsilon > 0$. While $\|\nabla f(x^k)\| \geq \epsilon$, do:

1. Form the local quadratic model $m_k(s)$ of $f(x^k + s)$.

2. Solve (approximately) the (TR) subproblem for s^k with $m_k(s^k) < f(x^k)$ ("sufficiently").

Compute $\rho_k := [f(x^k) - f(x^k + s^k)]/[f(x^k) - m_k(s^k)]$.

3. If $\rho_k \geq 0.9$, then [very successful step]
set $x^{k+1} := x^k + s^k$ and $\Delta_{k+1} := 2\Delta_k$.

Else if $\rho_k \geq 0.1$, then [successful step]
set $x^{k+1} := x^k + s^k$ and $\Delta_{k+1} := \Delta_k$.

Else [unsuccessful step]
set $x^{k+1} = x^k$ and $\Delta_{k+1} := \frac{1}{2}\Delta_k$.

4. Let $k := k + 1$. □

Proof of global convergence of the GTR method

Lemma 13: (Lower bound on TR radius) Let $f \in \mathcal{C}^2(\mathbb{R}^n)$ and ∇f be Lipschitz continuous on $\mathbb{R}^n$ with Lipschitz constant L. In GTR with Cauchy decrease $m_k(s^k) \leq m_k(s_c^k)$ for all k, suppose that

there exists $\epsilon > 0$ such that $\|\nabla f(x^k)\| \geq \epsilon$ for all k.

Then, there exists a constant $c \in (0,1)$ (independent of k) such that

$$\Delta_k \geq \frac{c}{L}\epsilon \quad \text{for all } k \geq 0.$$

Remarks:

(1) The proof of Lemma 13 relies on first showing that if $\Delta_k \leq \frac{2c}{L}\epsilon$, then iteration k is successful and $\Delta_{k+1} \geq \Delta_k$.

(2) If GTR takes finitely many successful iterations, then we can show that the last successful iterate has zero gradient. [$\Delta_k \to 0$ which contradicts L13 if gradient not zero.]

Proof of global convergence of the GTR method

Theorem 14: (At least one limit point is stationary) Let $f \in \mathcal{C}^2(\mathbb{R}^n)$ and and bounded below on $\mathbb{R}^n$. Let ∇f be Lipschitz continuous on $\mathbb{R}^n$ with Lipschitz constant L. Let $\{x^k\}$ be generated by the generic trust region (GTR) method, and let the computation of s^k be such that $m_k(s^k) \leq m_k(s_c^k)$ for all k. Then either there exists $k \geq 0$ such that $\nabla f(x^k) = 0$ or $\liminf_{k\to\infty} \|\nabla f(x^k)\| = 0$.

Proof of global convergence of the GTR method

Theorem 14: (At least one limit point is stationary) Let $f \in \mathcal{C}^2(\mathbb{R}^n)$ and and bounded below on $\mathbb{R}^n$. Let ∇f be Lipschitz continuous on $\mathbb{R}^n$ with Lipschitz constant L. Let $\{x^k\}$ be generated by the generic trust region (GTR) method, and let the computation of s^k be such that $m_k(s^k) \leq m_k(s_c^k)$ for all k. Then either there exists $k \geq 0$ such that $\nabla f(x^k) = 0$ or $\liminf_{k\to\infty} \|\nabla f(x^k)\| = 0$.

Proof of Theorem 14.

If there exists k such that $\nabla f(x^k) = 0$, then GTR terminates (this includes the case of having finitely many successful iterations).

Assume there exists $\epsilon > 0$ such that $\|\nabla f(x^k)\| \geq \epsilon$ for all k.

Proof of global convergence of the GTR method

Theorem 14: (At least one limit point is stationary) Let $f \in \mathcal{C}^2(\mathbb{R}^n)$ and and bounded below on $\mathbb{R}^n$. Let ∇f be Lipschitz continuous on $\mathbb{R}^n$ with Lipschitz constant L. Let $\{x^k\}$ be generated by the generic trust region (GTR) method, and let the computation of s^k be such that $m_k(s^k) \leq m_k(s_c^k)$ for all k. Then either there exists $k \geq 0$ such that $\nabla f(x^k) = 0$ or $\liminf_{k\to\infty} \|\nabla f(x^k)\| = 0$.

Proof of Theorem 14.

If there exists k such that $\nabla f(x^k) = 0$, then GTR terminates (this includes the case of having finitely many successful iterations).

Assume there exists $\epsilon > 0$ such that $\|\nabla f(x^k)\| \geq \epsilon$ for all k. Then using that there are infinitely many successful iterations $k \in \mathcal{S}$, and definition GTR scheme (namely, of ρ_k), we obtain

Proof of global convergence of the GTR method

Proof of Theorem 14 (continued).

$$\begin{aligned} f(x^k) - f(x^{k+1}) &\geq 0.1(f(x^k) - m_k(s^k)) \\ &\geq \tfrac{0.1}{2}\|\nabla f(x^k)\| \min\left\{\tfrac{\|\nabla f(x^k)\|}{\|\nabla^2 f(x^k)\|}, \Delta_k\right\} \end{aligned}$$

for all $k \in \mathcal{S}$, where we also used Lemma 12.

Proof of global convergence of the GTR method

Proof of Theorem 14 (continued).

$$\begin{aligned} f(x^k) - f(x^{k+1}) &\geq 0.1(f(x^k) - m_k(s^k)) \\ &\geq \tfrac{0.1}{2} \|\nabla f(x^k)\| \min\left\{\tfrac{\|\nabla f(x^k)\|}{\|\nabla^2 f(x^k)\|}, \Delta_k\right\} \end{aligned}$$

for all $k \in \mathcal{S}$, where we also used Lemma 12.

∇f Lipschitz cont. with Lips const $L \Longrightarrow \|\nabla^2 f(x)\| \leq L \ \forall x$.[Pb6,Sh2]

Proof of global convergence of the GTR method

Proof of Theorem 14 (continued).

$$\begin{aligned} f(x^k) - f(x^{k+1}) &\geq 0.1(f(x^k) - m_k(s^k)) \\ &\geq \frac{0.1}{2}\|\nabla f(x^k)\| \min\left\{\frac{\|\nabla f(x^k)\|}{\|\nabla^2 f(x^k)\|}, \Delta_k\right\} \end{aligned}$$

for all $k \in \mathcal{S}$, where we also used Lemma 12.

∇f Lipschitz cont. with Lips const $L \Longrightarrow \|\nabla^2 f(x)\| \leq L \ \forall x$.[Pb6,Sh2]

Thus since $\|\nabla f(x^k)\| \geq \epsilon$ for all k, we have for all $k \in \mathcal{S}$ that

$$f(x^k) - f(x^{k+1}) \geq 0.05\epsilon \min\left\{\frac{\epsilon}{L}, \Delta_k\right\} \geq 0.05\epsilon \min\left\{\frac{\epsilon}{L}, \frac{c}{L}\epsilon\right\},$$

where we also used Lemma 13.

Thus, using $c \in (0,1)$

for all $k \in \mathcal{S}$: $f(x^k) - f(x^{k+1}) \geq \frac{0.05c}{L}\epsilon^2$. (*)

Proof of global convergence of the GTR method

Proof of Theorem 14 (continued).

Since $f(x^k) \geq f_{\text{low}}$ for all k, we deduce for all $k \geq 0$,

$$\begin{aligned} f(x^0) - f_{\text{low}} &\geq f(x^0) - f(x^k) \\ &= \sum_{i=0}^{k-1} (f(x^i) - f(x^{i+1})) = \sum_{i \in \mathcal{S}}^{k-1} (f(x^i) - f(x^{i+1})) \end{aligned}$$

where we used $f(x^k) = f(x^{k+1})$ on all unsuccessful k.
Let $k \to \infty$. Then

$$\begin{aligned} f(x^0) - f_{\text{low}} &\geq \sum_{i=0}^{\infty} (f(x^i) - f(x^{i+1})) \\ &= \sum_{i \in \mathcal{S}} (f(x^i) - f(x^{i+1})) \geq |\mathcal{S}| \frac{0.05c}{L} \epsilon^2 \quad (**) \end{aligned}$$

we used (*) and $|\mathcal{S}| =$ no. of successful iterations. But LHS of (**) is finite while RHS of (**) is infinite since $|\mathcal{S}| = \infty$. Thus there must exist k such that $\|\nabla f(x^k)\| < \epsilon$.$\square$