Trust region methods (L7 and 8): Complete proof of global convergence (optional and non-examinable)

Coralia Cartis, Mathematical Institute, University of Oxford

C6.2/B2: Continuous Optimization

Global convergence of the GTR method

Theorem 11 (GTR global convergence)

Let $f \in \mathcal{C}^2(\mathbb{R}^n)$ and bounded below on $\mathbb{R}^n$. Let ∇f be Lipschitz continuous on $\mathbb{R}^n$ with Lipschitz constant $L \geq 1^{(*)}$. 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

$$\text{there exists } k \geq 0 \text{ such that } \nabla f(x^k) = 0$$

or

$$\lim_{k\to\infty} \|\nabla f(x^k)\| = 0.$$

[(*) $L \geq 1$ for convenience, to ease calculations.]

We (only) prove $\liminf_{k\to\infty} \|\nabla f(x^k)\| = 0$ (which also implies finite termination of GTR) next.

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

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

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

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

$$f(x^k) - m_k(s_c^k) \geq \Delta_k \|\nabla f(x^k)\|.$$

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 $\alpha_c^k h^k \leq \|\nabla f(x^k)\|^2$ and

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

Proof of global convergence of the GTR method

Proof of Lemma 12 (continued).

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

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

Finally, 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,

$$\begin{aligned} \frac{\|\nabla f(x^k)\|^4}{2h^k} &= \frac{\|\nabla f(x^k)\|^4}{2(\nabla f(x^k))^T\nabla^2 f(x^k)\nabla f(x^k)} \\ &\geq \frac{\|\nabla f(x^k)\|^2}{2\|\nabla^2 f(x^k)\|} \geq \frac{\|\nabla f(x^k)\|^2}{2(1+\|\nabla^2 f(x^k)\|)}(*). \end{aligned}$$

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

[(*) '$+1$' is only needed to cover the case $H^k = 0$.]

Proof of global convergence of the GTR method

Lemma 13: (Model error bound) Let $f \in \mathcal{C}^2(\mathbb{R}^n)$ and ∇f be Lipschitz continuous on $\mathbb{R}^n$ with Lipschitz constant L. Then in GTR, for all $k \geq 0$, we have $|f(x^k + s^k) - m_k(s^k)| \leq L\Delta_k^2$.

Proof of Lemma 13. Mean-value theorem gives

$$f(x^k + s^k) = f(x^k) + (s^k)^T \nabla f(x^k) + \frac{1}{2}(s^k)^T \nabla^2 f(\xi^k) s^k$$

for some ξ^k on line segment $[x^k, x^k + s^k]$. Then the definition of $m_k(s) = f(x^k) + s^T \nabla f(x^k) + \frac{1}{2} s^T \nabla^2 f(x^k) s$ gives

$$\begin{aligned}
|f(x^k + s^k) - m_k(s^k)| &\leq \tfrac{1}{2}|(s^k)^T \nabla^2 f(\xi^k) s^k - (s^k)^T \nabla^2 f(x^k) s^k| \\
&\leq \tfrac{1}{2}|(s^k)^T \nabla^2 f(\xi^k) s^k| + \tfrac{1}{2}|(s^k)^T \nabla^2 f(x^k) s^k| \\
&\leq \tfrac{1}{2}[\|\nabla^2 f(\xi^k)\| + \|\nabla^2 f(x^k)\|] \cdot \|s^k\|^2 \leq L\|s^k\|^2 \leq L\Delta_k^2,
\end{aligned}$$

where in the penultimate inequality we used that (∇f Lipschitz continuous with const. L) $\iff$ ($\|\nabla^2 f\|$ uniformly bounded above by L), and in the last inequality we used that $\|s^k\| \leq \Delta_k$. $\square$

Proof of global convergence of the GTR method

Lemma 14: (Successful iterations) Let $f \in \mathcal{C}^2(\mathbb{R}^n)$ and ∇f be Lipschitz continuous on $\mathbb{R}^n$ with Lipschitz constant $L \geq 1$. In GTR with Cauchy decrease $m_k(s^k) \leq m_k(s_c^k)$ for all k, suppose that $\nabla f(x^k) \neq 0$ and

$$\Delta_k \leq \frac{0.45}{L}\|\nabla f(x^k)\|. \qquad (1)$$

Then iteration k is successful and $\Delta_{k+1} \geq \Delta_k$.

Proof of Lemma 14. ∇f Lipschitz continuous on $\mathbb{R}^n$ with Lipschitz constant $L \geq 1 \implies 1 + \|\nabla^2 f(x)\| \leq 2L$ and so from (1), we deduce $\Delta_k \leq \dfrac{\|\nabla f(x^k)\|}{1+\|\nabla^2 f(x^k)\|}$. Lemma 12 now gives that

$$f(x^k) - m_k(s^k) \geq \frac{1}{2}\|\nabla f(x^k)\|\Delta_k > 0$$

and Lemma 13 that $|f(x^k+s^k) - m_k(s^k)| \leq L\Delta_k^2$. We evaluate

$|\rho_k - 1| = \frac{|f(x^k+s^k)-m_k(s^k)|}{f(x^k)-m_k(s^k)} \leq \frac{2L\Delta_k^2}{\Delta_k\|\nabla f(x^k)\|} = \frac{2L\Delta_k}{\|\nabla f(x^k)\|} \leq 0.9 \Rightarrow \rho_k \geq 0.1.$ □

Proof of global convergence of the GTR method

Lemma 15: (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 $L \geq 1$. 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

$$\Delta_k \geq \frac{0.45}{2L}\epsilon \quad \text{for all } k \geq 0.$$

Proof of Lemma 15. Assume the contrary: k is the first iteration such that $\Delta_{k+1} < \frac{0.45}{2L}\epsilon$. Then k unsuccessful and $\Delta_{k+1} = \frac{\Delta_k}{2}$. Thus $\Delta_k = 2\Delta_{k+1} < \frac{0.45}{L}\epsilon \leq \frac{0.45}{L}\|\nabla f(x^k)\|$ and so by Lemma 14, k must be successful, contradiction. $\square$

Proof of global convergence of the GTR method

Theorem 16: (The case of finitely many successful iterations) Let $f \in \mathcal{C}^2(\mathbb{R}^n)$ and ∇f be Lipschitz continuous on $\mathbb{R}^n$ with Lipschitz constant $L \geq 1$. In GTR with Cauchy decrease $m_k(s^k) \leq m_k(s_c^k)$ for all k, suppose that there are finitely many successful iterations that occur. Then $x^k = x_*$ for all k sufficiently large and $\nabla f(x_*) = 0$.

Proof of Theorem 16. Let k_o be the last successful iteration. Then GTR implies $x^k = x_*$ for all $k \geq k_o + 1$. As all remaining iterations are unsuccessful, $\Delta_{k+1} = \frac{1}{2}\Delta_k$ for all $k \geq k_o + 1$ and so $\Delta_k \longrightarrow 0$ as $k \to \infty$. If $\nabla f(x^{k_o+1}) \neq 0$, then let $\epsilon = \|\nabla f(x^{k_o+1})\|$ in Lemma 15, which implies that Δ_k is bounded away from zero; contradiction. Thus $\nabla f(x^{k_o+1}) = 0$ and so $\nabla f(x^k) = \nabla f(x_*) = 0$ for all $k \geq k_o + 1$. $\Box$

Proof of global convergence of the GTR method

Theorem 17: (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 $L \geq 1$. 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 17. If there exists k such that $\nabla f(x^k) = 0$, then GTR terminates. Assume there exists $\epsilon > 0$ such that $\|\nabla f(x^k)\| \geq \epsilon$ for all k. Then Th 16 implies that there are infinitely many successful iterations $k \in \mathcal{S}$, and from GTR/ρ_k,

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

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

Proof of global convergence of the GTR method

Proof of Theorem 17 (continued).

∇f Lipschitz cont. with Lips const $L \geq 1 \Longrightarrow \|\nabla^2 f(x)\| \leq L \; \forall x$.
Thus since $\|\nabla f(x^k)\| \geq \epsilon$ for all k, we have for all $k \in \mathcal{S}$ that

$$f(x^k) - f(x^{k+1}) \geq 0.05\epsilon \min\left\{\frac{\epsilon}{2L}, \Delta_k\right\} \geq 0.05\epsilon \min\left\{\frac{\epsilon}{2L}, \frac{0.45}{2L}\epsilon\right\},$$

where we also used Lemma 15. Thus

$$\text{for all } k \in \mathcal{S}: \quad f(x^k) - f(x^{k+1}) \geq \frac{0.01}{2L}\epsilon^2. \quad (*)$$

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

$$f(x^0) - f_{\text{low}} \geq f(x^0) - \lim_{k\to\infty} f(x^k) \geq \sum_{i=0}^{\infty}(f(x^i) - f(x^{i+1}))$$
$$= \sum_{i\in\mathcal{S}}(f(x^i) - f(x^{i+1})) \geq |\mathcal{S}|\frac{0.01}{2L}\epsilon^2 \quad (**)$$

where in '=' we used $f(x^k) = f(x^{k+1})$ on all unsuccessful k, and in the last '$\geq$', 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 $\|\nabla f(x^k)\| < \epsilon$.$\Box$