Lectures 7 and 8: Trust-region methods for unconstrained optimization

Coralia Cartis, Mathematical Institute, University of Oxford

C6.2/B2: Continuous Optimization

Linesearch versus trust-region methods

(UP): minimize $f(x)$ subject to $x \in \mathbb{R}^n$.

Linesearch methods: 'liberal' in the choice of search direction, keeping bad behaviour in control by choice of α^k.

- choose descent direction s^k,
- compute stepsize α^k to reduce $f(x^k + \alpha s^k)$,
- update $x^{k+1} := x^k + \alpha^k s^k$.

Trust region (TR) methods: 'conservative' in the choice of search direction, so that a full stepsize along it may really reduce the objective.

- pick direction s^k to reduce a "local model" of $f(x^k + s^k)$,
- accept $x^{k+1} := x^k + s^k$ if decrease in the model is also achieved by $f(x^k + s^k)$,
- else set $x^{k+1} := x^k$ and "refine" the model.

Trust-region models for unconstrained problems

Approximate $f(x^k + s)$ by:

- linear model $\quad l_k(s) := f(x^k) + s^\top \nabla f(x^k) \qquad$ or
- quadratic model

$$q_k(s) := f(x^k) + s^\top \nabla f(x^k) + \frac{1}{2} s^\top \nabla^2 f(x^k) s.$$

Impediments:

models may not resemble $f(x^k + s)$ when s is large,

models may be unbounded from below,

* $l_k(s)$ always unbounded below (unless $\nabla f(x^k) = 0$)
* $q_k(s)$ is always unbounded below if $\nabla^2 f(x^k)$ is negative definite or indefinite, and sometimes if $\nabla^2 f(x^k)$ is positive semidefinite.

Trust region models and subproblem

Prevent bad approximations by trusting the model only in a trust region, defined by the trust region constraint

$$\|s\| \leq \Delta_k, \tag{R}$$

for some "appropriate" radius $\Delta_k > 0$.

The constraint (R) also prevents l_k, q_k from unboundedness!

$\Longrightarrow$ the trust region subproblem

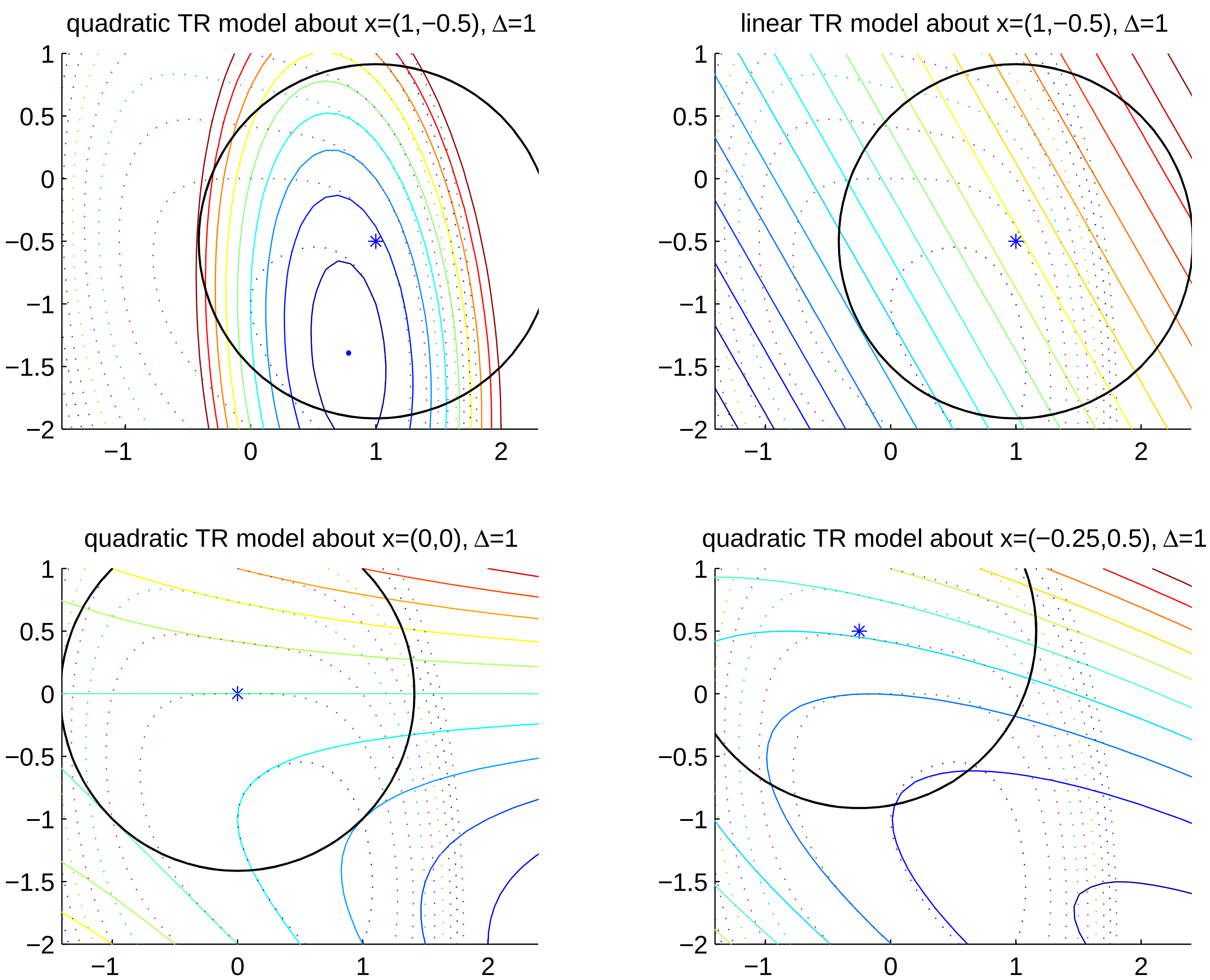
$$\min_{s\in\mathbb{R}^n} m_k(s) \quad \text{subject to} \quad \|s\| \leq \Delta_k, \tag{TR}$$

where $m_k := l_k,\ k \geq 0$, or $m_k := q_k,\ k \geq 0$.

- From now on, $m_k := q_k$.

(TR) easier to solve than (P). May even solve (TR) only approximately.

Trust region models and subproblem - an example



Trust-region models of $f(x) = x_1^4 + x_1 x_2 + (1 + x_2)^2$.

Generic trust-region method

Let s^k be a(n approximate) solution of (TR). Then

- predicted model decrease:
  $$m_k(0) - m_k(s^k) = f(x^k) - m_k(s^k).$$
- actual function decrease: $f(x^k) - f(x^k + s^k)$.

The trust region radius Δ_k is chosen based on the value of

$$\rho_k := \frac{f(x^k) - f(x^k + s^k)}{f(x^k) - m_k(s^k)}.$$

If ρ_k is not too smaller than 1, $x^{k+1} := x^k + s^k$, $\Delta_{k+1} \geq \Delta_k$.

If ρ_k close to or ≥ 1, Δ_k is increased.

If $\rho_k \ll 1$, $x^{k+1} = x^k$ and Δ_k is reduced.

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 $k := k + 1$. $\square$

Trust-region methods

- Other sensible values of the parameters of the GTR are possible.

"Solving" the (TR) subproblem

$$\min_{s \in \mathbb{R}^n} m_k(s) \quad \text{subject to} \quad \|s\| \leq \Delta_k, \tag{TR}$$

... exactly or even approximately may imply work.

Want "minimal" condition of "sufficient decrease" in the model that ensures global convergence of the TR method (the Cauchy cond.). In practice, we (usually) do much better than this condition!

Example of applying a trust-region method: [Sartenaer, 2008].

- approximate solution of (TR) subproblem: better than Cauchy, but not exact.
- notation: $\Delta f / \Delta m_k \equiv \rho_k$.

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

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

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

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

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

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

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

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

there exists $k \geq 0$ 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 $\alpha_c^k > 0$.
  - $h^k > 0$: $\alpha_{\min} := \dfrac{\|\nabla f(x^k)\|^2}{h^k} = \arg\min_{\alpha>0}\phi(\alpha)$.
    $\Longrightarrow \quad \alpha_c^k = \min(\alpha_{\min}, \overline{\alpha})$.
  - $h^k \leq 0$: $\phi(\alpha)$ 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

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

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

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

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

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 $L \geq 1 \Longrightarrow 1 + \|\nabla^2 f(x)\| \leq 2L$ 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. □

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 $k \geq k_o + 1$. $\square$

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,

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

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

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

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

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

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 $\|\nabla f(x^k)\| < \epsilon$.$\Box$

Solving the (TR) subproblem

On each TR iteration we compute or approximate the solution of

$$\min_{s\in\mathbb{R}^n} m_k(s) = f(x^k) + s^\top \nabla f(x^k) + \frac{1}{2} s^\top \nabla^2 f(x^k) s$$
$$\text{subject to } \|s\| \leq \Delta_k.$$

- also, s^k must satisfy the Cauchy condition $m_k(s^k) \leq m_k(s_c^k)$, where $s_c^k := -\alpha_c^k \nabla f(x^k)$, with

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

[Cauchy condition ensures global convergence]

• solve (TR) exactly (i.e., compute global minimizer of TR) $\Longrightarrow$ TR akin to Newton-like method.

• solve (TR) approximately (i.e., an approximate global minimizer) $\Longrightarrow$ large-scale problems.

Solving the (TR) subproblem exactly

For $h \in \mathbb{R}$, $\Delta > 0$, $g \in \mathbb{R}^n$, H $n \times n$ symm. matrix, consider

$$\min_{s \in \mathbb{R}^n} m(s) := h + s^\top g + \frac{1}{2} s^\top H s, \text{ s. t. } \|s\| \leq \Delta. \qquad \text{(TR)}$$

Characterization result for the solution of (TR):

Theorem 18

Any global minimizer s^* of (TR) satisfies the equation

$$(H + \lambda^* I) s^* = -g,$$

where $H + \lambda^* I$ is positive semidefinite, $\lambda^* \geq 0$,

$$\lambda^* (\|s^*\| - \Delta) = 0 \quad \text{and} \quad \|s^*\| \leq \Delta.$$

If $H + \lambda^* I$ is positive definite, then s^* is unique.

- The above Theorem gives necessary and sufficient global optimality conditions for a nonconvex optimization problem!

Solving the (TR) subproblem exactly

Computing the global solution s^* of (TR):

Case 1. If H is positive definite and $Hs = -g$ satisfies $\|s\| \leq \Delta$ $\Longrightarrow s^* := s$ (unique), $\lambda^* := 0$ (by Theorem 18).

Case 2. If H is positive definite but $\|s\| > \Delta$,
or H is not positive definite, Theorem 18 implies s^* satisfies

$$(H + \lambda I)s = -g, \quad \|s\| = \Delta, \tag{$*$}$$

for some $\lambda \geq \max\{0, -\lambda_{\min}(H)\} := \underline{\lambda}$.

Let $s(\lambda) = -(H + \lambda I)^{-1} g$, for any $\lambda > \underline{\lambda}$. Then $s^* = s(\lambda^*)$ where $\lambda^* \geq \underline{\lambda}$ solution of

$$\|s(\lambda)\| = \Delta, \quad \lambda \geq \underline{\lambda}.$$

$\longrightarrow$ nonlinear equation in one variable λ. Use Newton's method to solve it. We discuss the system (*) in detail next.

Solving the (TR) subproblem exactly ...

$$(H + \lambda I)s = -g, \quad s^\top s = \Delta^2. \qquad (*)$$

H symmetric $\Longrightarrow$ spectral decomposition: $H = U^\top \Lambda U$, with U orthonormal matrix of the eigenvectors of H and Λ diagonal mat. of eigenvalues of H, $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$; $\lambda_1 = \lambda_{\min}(H)$

Th. 18 $\Longrightarrow$ $H + \lambda I = U^\top(\Lambda + \lambda I)U$ positive semidefinite $\Longrightarrow$ $\lambda_1 + \lambda \geq 0 \Longrightarrow \lambda \geq -\lambda_1 \quad \Longrightarrow \quad \lambda \geq \max\{0, -\lambda_1\}$.

$\lambda \longrightarrow s(\lambda) := -(H + \lambda I)^{-1} g$, provided $H + \lambda I$ nonsingular.

$\psi(\lambda) := \|s(\lambda)\|^2 = \|U^\top(\Lambda + \lambda I)^{-1} U g\|^2 = g^\top U^\top (\Lambda + \lambda I)^{-2} U g$

- $g = U^\top \gamma$, for some $\gamma = (\gamma_1, \ldots, \gamma_n) \in \mathbb{R}^n$. As $UU^\top = U^\top U = I$,

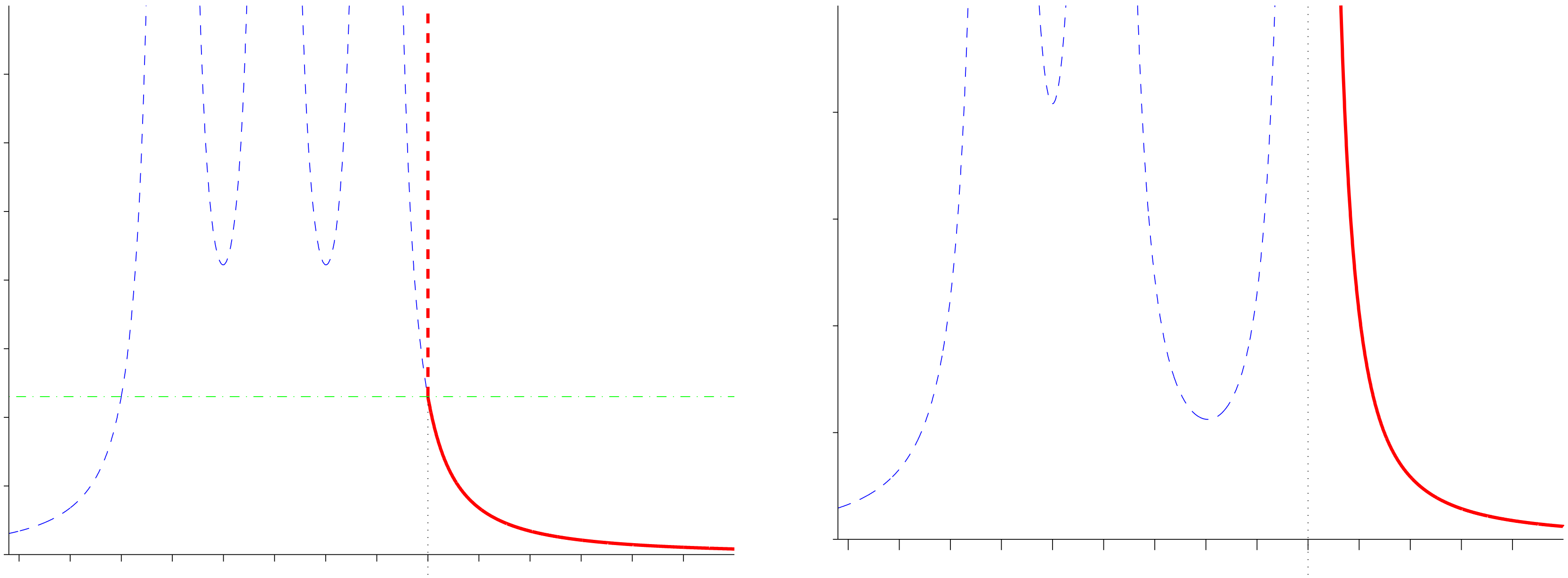
$$\psi(\lambda) = \gamma^\top (\Lambda + \lambda I)^{-2} \gamma = \sum_{i=1}^{n} \frac{\gamma_i^2}{(\lambda + \lambda_i)^2} \overset{(*)}{=} \Delta^2.$$

The secular equation

Consider

$$\psi(\lambda) := \|s(\lambda)\|^2 = \Delta^2$$

for $\lambda \in (\max\{0, -\lambda_1\}, \infty)$. [see Pb Sheet 4]



'Easy' cases: Plots of λ vs. $\psi(\lambda)$; $H \succ 0$ (LHS) and H indef (RHS).

The secular equation

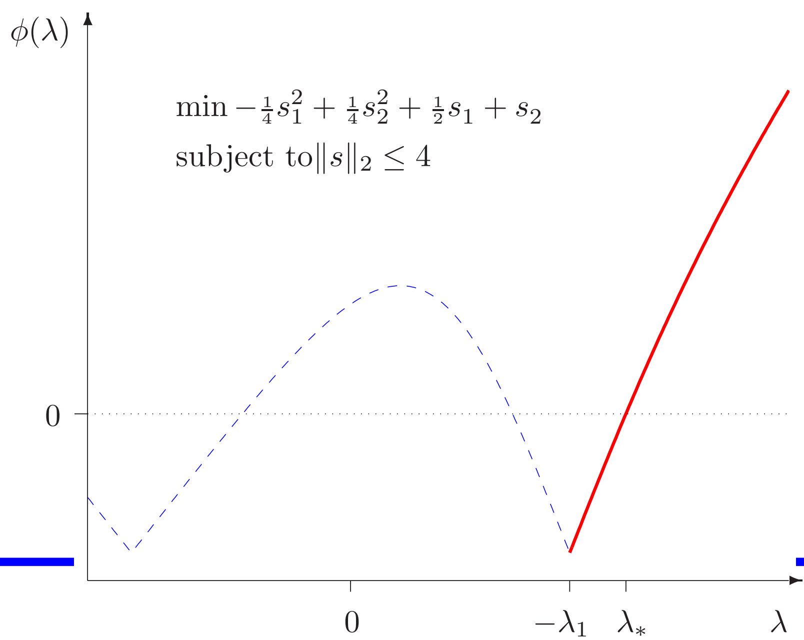
DON'T solve $\psi(\lambda) := \|s(\lambda)\|^2 = \Delta^2$.

Solve instead the secular equation

$$\phi(\lambda) := \frac{1}{\|s(\lambda)\|} - \frac{1}{\Delta} = 0 \text{ for } \lambda \in (\max\{0, -\lambda_1\}, \infty). \qquad (\dagger)$$

- ϕ has no poles; it is analytic on $(-\lambda_1, \infty)$
  $\Longrightarrow$ ideal for Newton's mthd (exc. in the 'hard' case).

[globally convergent and locally quadratic if $\lambda^0 \in [-\lambda_1, \lambda_*]$; else safeguard with linesearch]



Solving the (TR) subproblem for large-scale problems

• Newton's mthd for (†): Cholesky factorization of $H + \lambda I$ for various $\lambda \longrightarrow$ expensive or impossible for large problems.

No computation of the complete eigenvalue decomposition of H!

Solving the large-scale (TR) subproblem:

• Use iterative methods to approximate the global minimizer of (TR).

Use the Cauchy point (i.e. steepest descent): impractical.

Use conjugate-gradient or Lanczos method (as the first step is a steepest descent, and thus our requirement of "sufficient decrease" in m_k will be satisfied).

Nonlinear least-squares/inverse problems

- $r : \mathbb{R}^n \rightarrow \mathbb{R}^m$ with $m \geq n$; r smooth.

  $$\min_{x \in \mathbb{R}^n} f(x) := \frac{1}{2} \sum_{j=1}^{m} [r_j(x)]^2 = \frac{1}{2} \|r(x)\|^2. \qquad \text{(NLS)}$$

The Levenberg-Marquardt method: replace linesearch in Gauss-Newton with trust-region

$$\Longrightarrow \quad \min_{s \in \mathbb{R}^n} \frac{1}{2} \|J(x^k)s + r(x^k)\|^2 \text{ subject to } \|s\| \leq \Delta_k.$$

- useful when $J(x^k)$ is rank-deficient (i.e., not full-rank); overcomes weakness of Gauss-Newton.

- s^k solves TR subproblem iff $\exists \lambda^k \geq 0$ such that

  $$(J(x^k)^T J(x^k) + \lambda^k I)s^k = -J(x^k)^T r(x^k)$$

  and $\lambda^k(\|s^k\| - \Delta_k) = 0$.

Linesearch vs. trust-region methods

Quasi-Newton methods/approximate derivatives also possible in the trust-region framework; no need for positive definite updates for the Hessian! Replace $\nabla^2 f(x^k)$ with approximation B^k in the quadratic local model $m_k(s)$.

Conclusions: state-of-the-art software for unconstrained problems implements linesearch or TR methods; both approaches have been made competitive (more heuristics needed by linesearch methods to deal with negative curvature). Choosing between the two is mostly a matter of “taste”.

Information on existing software can be found at the NEOS Center: `http://www.neos-guide.org`

$\longrightarrow$ look under Optimization Guide and Optimization Tree, etc.

State-of-the-art NLO software: KNITRO, IPOPT, GALAHAD,...