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

Coralia Cartis, Mathematical Institute, University of Oxford

C6.2/B2: Continuous Optimization

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

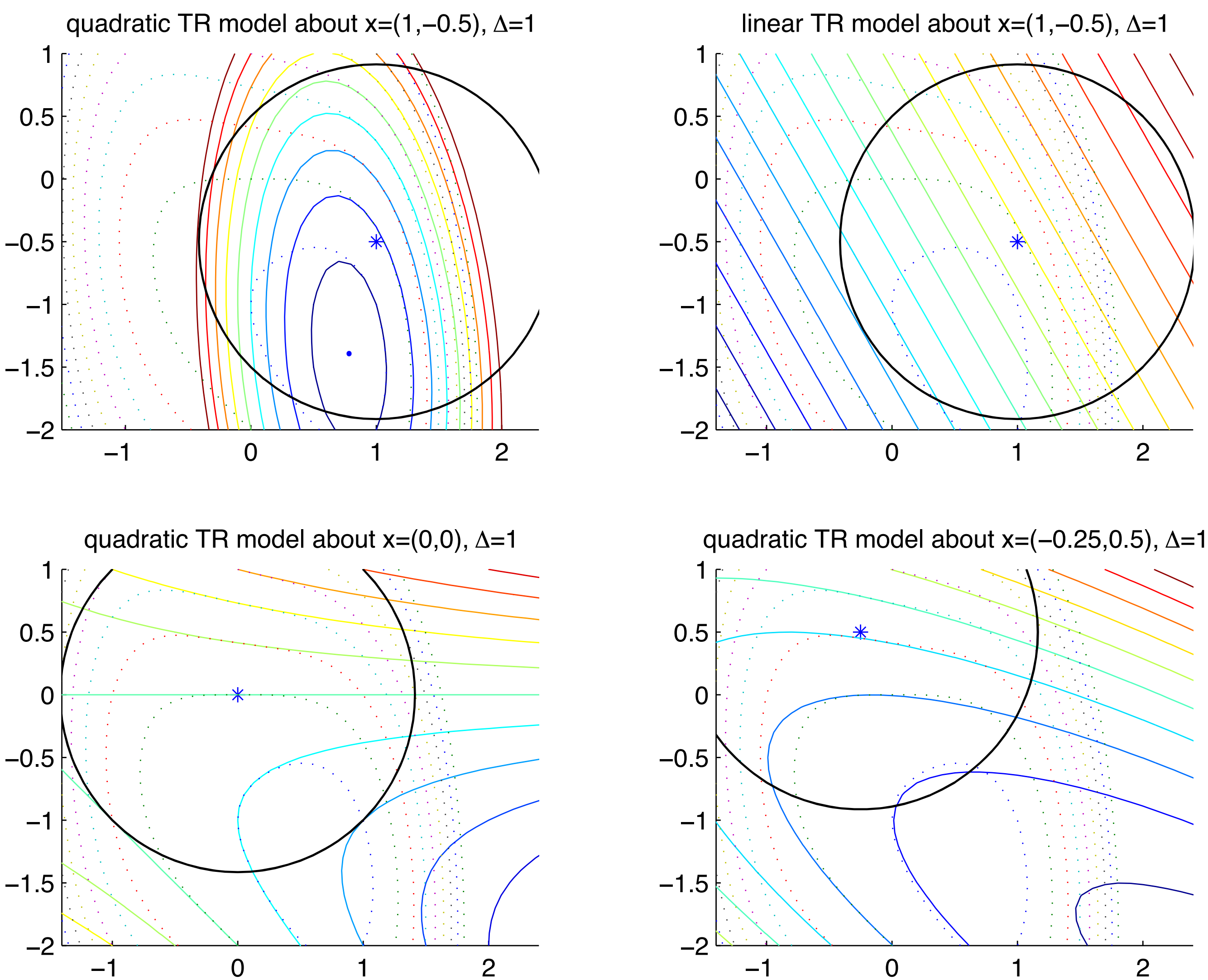
$$\alpha_c^k := \arg\min_{\alpha>0} m_k(-\alpha \nabla f(x^k)) \text{ subject to } \|\alpha \nabla f(x^k)\| \leq \Delta_k.$$

[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.

Trust region models and subproblem - an example



Trust-region models of $f(x) = x_1^4 + x_1x_2 + (1 + x_2)^2$.

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):

<u>Theorem 15</u>

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 15).

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 15).

Case 2. If H is positive definite but $\|s\| > \Delta$,
or H is not positive definite, Theorem 15 implies s^* satisfies

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

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

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 15).

Case 2. If H is positive definite but $\|s\| > \Delta$,
or H is not positive definite, Theorem 15 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 ...

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

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, $\lambda_1 \leq \lambda_2 \leq ... \leq \lambda_n$; $\lambda_1 = \lambda_{\min}(H)$

Solving the (TR) subproblem exactly ...

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

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. 15 $\Longrightarrow$ $H + \lambda I = U^\top(\Lambda + \lambda I)U$ positive semidefinite $\Longrightarrow$ $\lambda_1 + \lambda \geq 0 \Longrightarrow \lambda \geq -\lambda_1 \quad \Longrightarrow \quad \lambda \geq \max\{0, -\lambda_1\}$.

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, $\lambda_1 \leq \lambda_2 \leq ... \leq \lambda_n$; $\lambda_1 = \lambda_{\min}(H)$

Th. 15 $\Longrightarrow$ $H + \lambda I = U^\top(\Lambda + \lambda I)U$ 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 $UU^\top = U^\top U = I$,

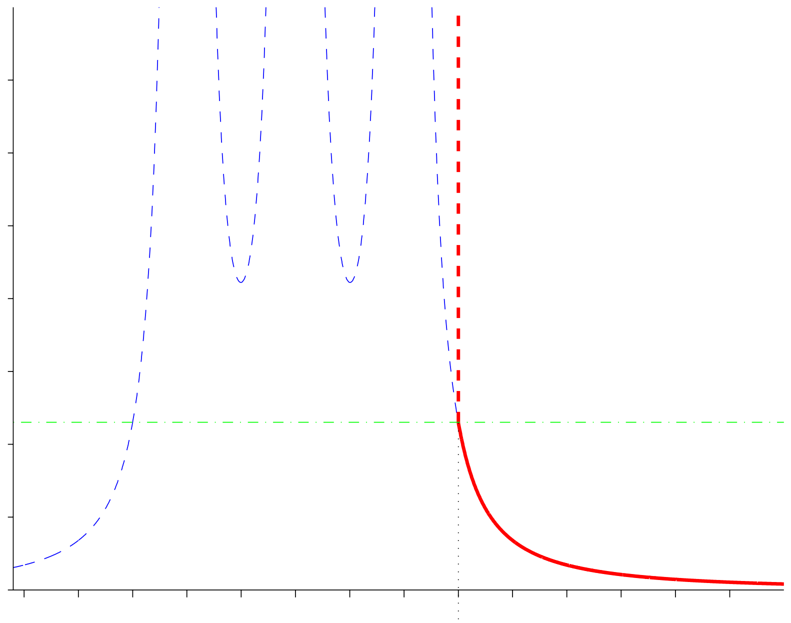
$$\psi(\lambda) = \gamma^\top (\Lambda + \lambda I)^{-2} \gamma = \sum_{i=1}^{n} \frac{\gamma_i^2}{(\lambda + \lambda_i)^2} \overset{(*)}{=} \Delta^2.$$

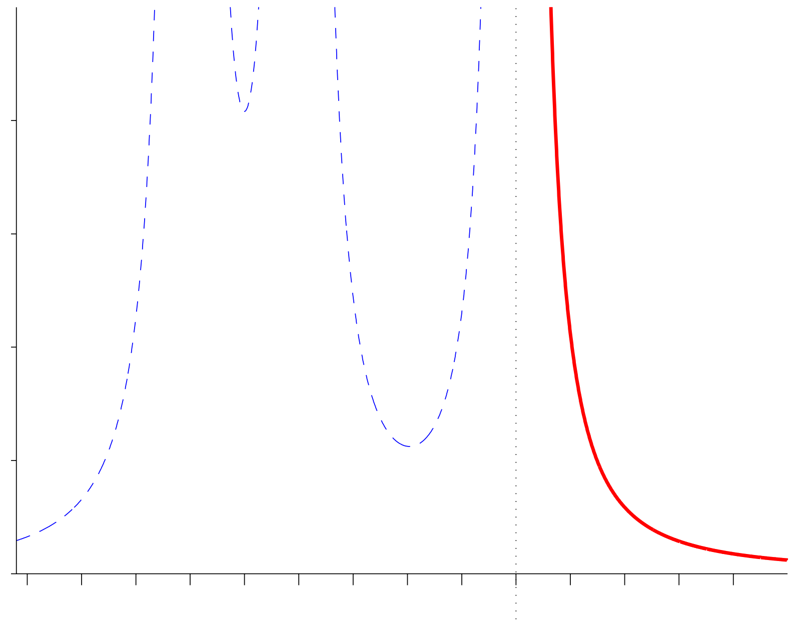
The secular equation

Consider

$$\psi(\lambda) := \|s(\lambda)\|^2 = \sum_{i=1}^{n} \frac{\gamma_i^2}{(\lambda + \lambda_i)^2} = \Delta^2$$

for $\lambda \in (\max\{0, -\lambda_1\}, \infty)$. [see Pb Sheet 3]





'Easy' cases: Plots of λ vs. $\psi(\lambda)$; $H \succ 0$ (LHS) and H indef (RHS). $n = 3$ in the plots.

The secular equation

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

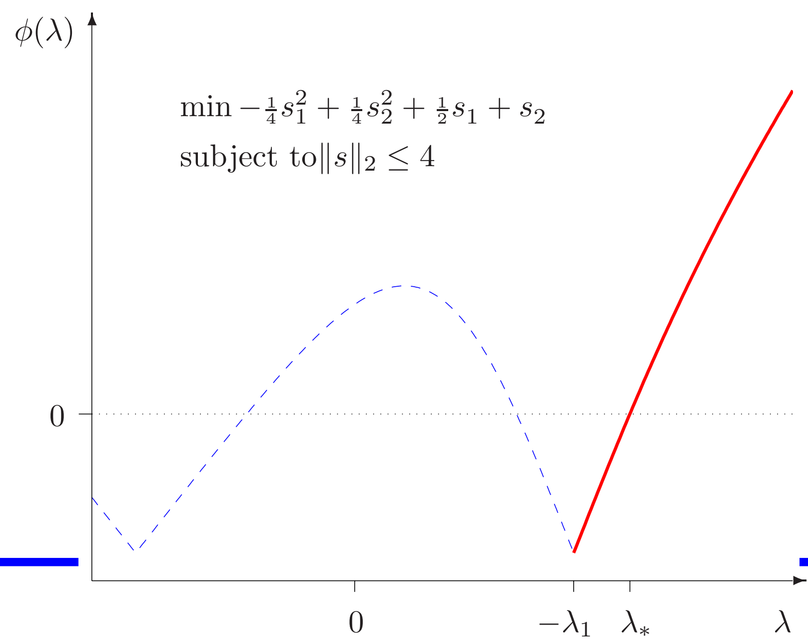
Solve instead the secular equation

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

- 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).

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,...