4.1 TRUST REGION FUNDAMENTALS

CHICAGO

Trust Region Idea

- Notations

$$f^k = f\left(x^k\right) \quad \nabla f^k = \nabla f\left(x^k\right)$$

- Quadratic Model

$$m_k(p) = f^k + p^T g^k + \frac{1}{2} p^T B^k p$$

- Order of Quadratic Model (Taylor)

$$f\left(x^k + p\right) = f^k + p^T g^k + \frac{1}{2} p^T \nabla_{xx}^2 f\left(x^k + tp\right) p \quad t \in [0,1]$$

$$m_k(p) - f\left(x^k + p\right) = \begin{cases} O\left(\|p\|^2\right) \\ O\left(\|p\|^2\right) & B^k = \nabla_{xx}^2 f\left(x^k\right) \end{cases}$$

CHICAGO

Trust Region Subproblem

$$\min_{p \in R^n} \quad m_k(p)$$

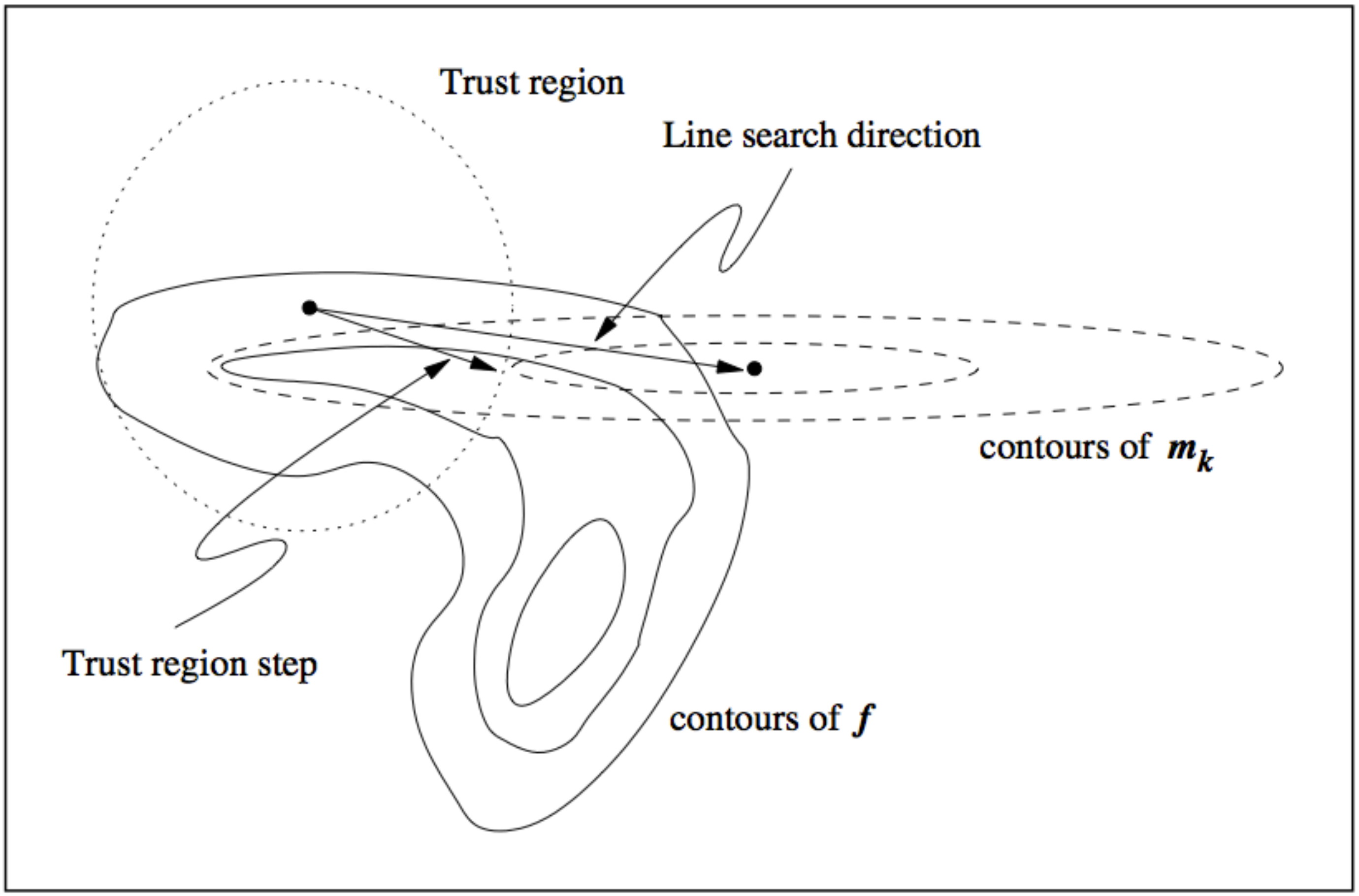
$$\text{subject to} \quad \|p\| \le \Delta^k$$

Called Trust Region
Constraint

- If $B^k \succ 0$ and $p^{*k} = \left(B^k\right)^{-1} g^k$; where $\|p^{*k}\| \le \Delta^k$ then p^k is the solution of the TR subproblem.
- But the interesting case lies in the opposite situation (since not, why would you need the TR in first place)?

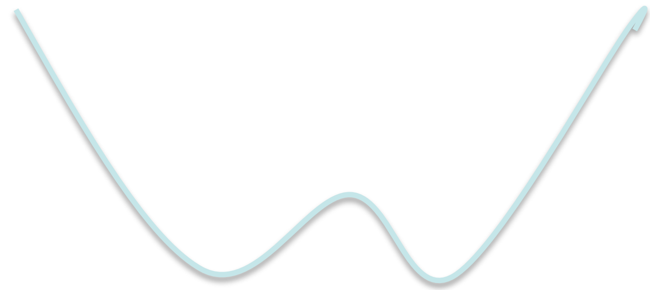
CHICAGO

Trust Region Geometric Intuition



Example

$$\min_x \left(x^2 - 1\right)^2$$



- Line search started at 0 cannot progress.
- How about the trust-region?

$$\min_d -2d^2; \quad |d| \le \Delta$$

- Either solution will escape the saddle point -- that is the principle of trust-region.

General approach

- How do we solve the TR subproblem?
- If $B^k \succ 0$ (or if we are not obsessed with stopping at saddle points) we use "dogleg" method. (LS, NLE). Most linear algebra is in computing

$$B^k d^{k,U} = -g^k$$

- If fear saddle points, we have to mess around with eigenvalues and eigenvectors – much harder problem.

CHICAGO

Trust Region Management: Parameters

- The quality of the reduction.

Actual Reduction

$$\rho^k = \frac{f\left(x^k\right) - f\left(x^k + p^k\right)}{m_k(0) - m_k\left(p^k\right)}$$

Predicted Reduction

- Define the acceptance ratio

$$\eta \in \left[0, \frac{1}{4}\right)$$

- Define the maximum TR size

$$\hat{\Delta}; \quad \Delta \in \left[0, \hat{\Delta}\right)$$

CHICAGO

TR management

Algorithm 4.1 (Trust Region).

Given $\hat{\Delta} > 0$, $\Delta_0 \in (0, \hat{\Delta})$, and $\eta \in \left[0, \frac{1}{4}\right)$:

for $k = 0, 1, 2, \ldots$

  Obtain p_k by (approximately) solving (4.3);

  Evaluate ρ_k from (4.4);

  if $\rho_k < \frac{1}{4}$

    $\Delta_{k+1} = \frac{1}{4}\Delta_k$

  else

    if $\rho_k > \frac{3}{4}$ and $\|p_k\| = \Delta_k$

      $\Delta_{k+1} = \min(2\Delta_k, \hat{\Delta})$

    else

      $\Delta_{k+1} = \Delta_k$;

  if $\rho_k > \eta$

    $x_{k+1} = x_k + p_k$

  else

    $x_{k+1} = x_k$;

I will ask you to code It with dogleg

CHICAGO

What if I cannot solve the TR exactly ?

- Since it is a hard problem.
- Will this destroy the "Global" convergence behavior?
- Idea: Accept a "sufficient" reduction.
- But, I have no Armijo (or Wolfe, Goldshtein criterion) …
- What do I do?
- Idea? Solve a simple TR problem that creates the yardstick for acceptance – the Cauchy point.

CHICAGO

4.2 THE CAUCHY POINT

The Cauchy Point

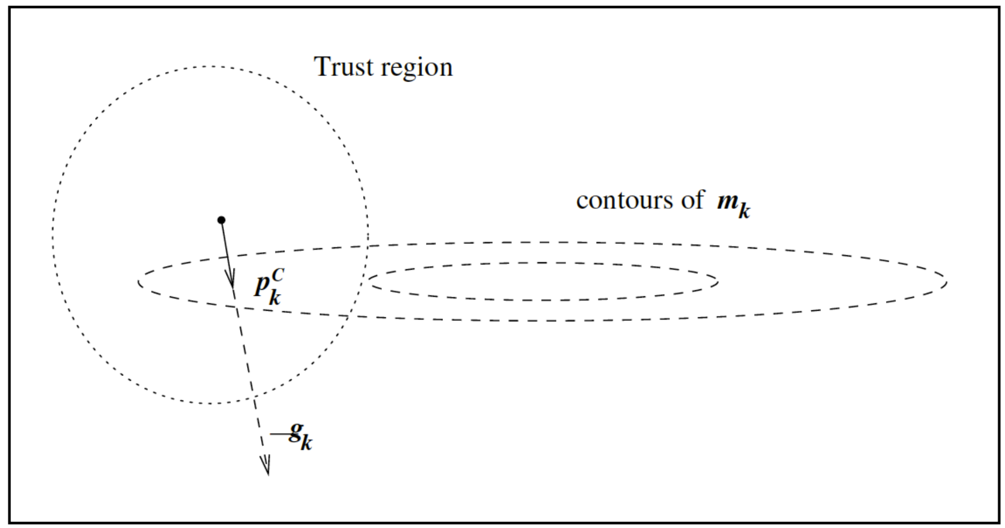
- What is an easy model to solve? Linear model
$$l_k(p) = f^k + g^{k,T} p$$
- Solve TR linear model
$$p^{k,s} = \arg\min_{p \in R^n \quad \|p\| \le \Delta^k} l_k(p)$$
- The Cauchy point.
$$\tau^k = \arg\min_{\tau \in R \quad \|\tau p^{k,s}\| \le \Delta^k} m_k(\tau p^{k,s})$$
$$p^{k,c} = \tau^k p^{k,s}; \quad x^{k,c} = x^k + p^{k,c}$$
- The reduction $m(0) - m(p^{k,c})$ becomes my yardstick; if trust region has at least this decrease, I can guarantee "global" convergence (reduction is $O(\|g^k\|^2)$)

Cauchy Point Solution

- First, solution of the linear problem is

$$p_k^s = -\frac{\Delta^k}{\|g^k\|} g^k$$

- Then, it immediately follows that



$$\tau_k = \begin{cases} 1 & g_k^T B_k g_k \le 0 \\ \min\left(\frac{\|g_k\|^3}{\left(g_k^T B_k g_k\right) \Delta_k}, 1 \right) & \text{otherwise} \end{cases}$$

CHICAGO

Dogleg Methods: Improve CP

- If Cauchy point is on the boundary I have a lot of decrease and I accept it (e.g if $g^{k,T} B_k g^k \leq 0;$)
- If Cauchy point is interior,

$$g^{k,T} B_k g^k > 0; \quad p^{k,c} = -\frac{\left\| g_k \right\|^2}{g^{k,T} B_k g^k} g^k$$

- Take now "Newton" step $p^B = -B_k^{-1} g^k$ (note, B need not be pd, all I need is nonsingular).

CHICAGO

Dogleg Method Continued

I will ask you to code it with TR

- Define dogleg path

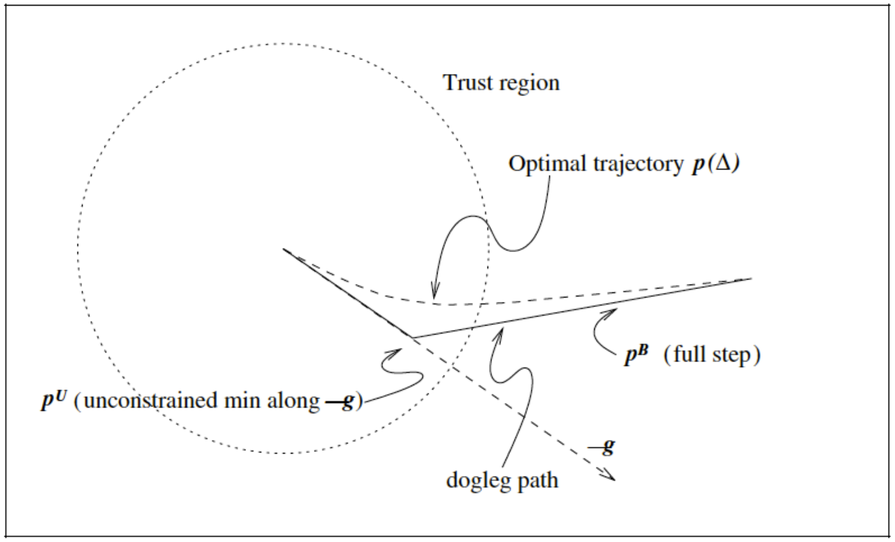
$$\tilde{p}(\tau)=\begin{cases} \tau p^{k,c} & \tau \le 1 \\ p^{k,c}+(\tau-1)\left(p^{B}-p^{k,c}\right) & 1\le \tau \le 2 \end{cases}$$

- The dogleg point:

$$\tilde{p}(\tau_D); \quad \tau_D = \arg\min_{\tau;\|\tilde{p}(\tau)\|\le \Delta_k} m_k\left(\tilde{p}(\tau)\right)$$

- It is obtained by solving 2 quadratics.
- Sufficiently close to the solution it allows me to choose the Newton step, $\tau = 2$ and thus quadratic convergence.

Dogleg Method: Theory



Lemma 4.2.

Let B be positive definite. Then

(i) $\|\tilde{p}(\tau)\|$ *is an increasing function of* τ, *and*

(ii) $m(\tilde{p}(\tau))$ *is a decreasing function of* τ.

Global Convergence of CP Methods

Lemma 4.3.

The Cauchy point p_k^C satisfies (4.20) with $c_1 = \frac{1}{2}$, that is,

$$m_k(0) - m_k(p_k^C) \geq \tfrac{1}{2}\|g_k\| \min\left(\Delta_k, \frac{\|g_k\|}{\|B_k\|}\right).$$

$$\|p_k\| \leq \gamma \Delta_k, \qquad \text{for some constant } \gamma \geq 1. \tag{4.25}$$

$$m_k(0) - m_k(p_k) \geq c_1\|g_k\| \min\left(\Delta_k, \frac{\|g_k\|}{\|B_k\|}\right), \tag{4.20}$$

Theorem 4.5.

Let $\eta = 0$ in Algorithm 4.1. Suppose that $\|B_k\| \leq \beta$ for some constant β, that f is bounded below on the level set S defined by (4.24) and Lipschitz continuously differentiable in the neighborhood $S(R_0)$ for some $R_0 > 0$, and that all approximate solutions of (4.3) satisfy the inequalities (4.20) and (4.25), for some positive constants c_1 and γ. We then have

$$\liminf_{k\to\infty} \|g_k\| = 0. \tag{4.26}$$

Numerical comparison between methods

- What is a fair comparison between methods?
- Probably : starting from same point 1) number of function evaluations and 2) number of linear systems (the rest depends too much on the hardware and software platform). I will ask you to do this.
- Trust region tends to use fewer function evaluations (the modern preferred metric;) than line search .
- Also dogleg does not force positive definite matrix, so it has fewer chances of stopping at a saddle point, (but it is not guaranteed either).

CHICAGO

4.3 GENERAL CASE: SOLVING THE ACTUAL TR PROBLEM (DOGLEG DOES NOT QUITE DO IT)

Trust Region Equation

Theorem 4.1.

The vector p^* is a global solution of the trust-region problem

$$\min_{p \in \mathbb{R}^n} m(p) = f + g^T p + \frac{1}{2} p^T B p, \quad \text{s.t. } \|p\| \le \Delta, \tag{4.7}$$

if and only if p^* is feasible and there is a scalar $\lambda \ge 0$ such that the following conditions are satisfied:

$$(B + \lambda I) p^* = -g, \tag{4.8a}$$

$$\lambda(\Delta - ||p^*||) = 0, \tag{4.8b}$$

$$(B + \lambda I) \quad \text{is positive semidefinite.} \tag{4.8c}$$

CHICAGO

Theory of Trust Region Problem

Global convergence away from saddle point

Theorem 4.8.

Suppose that the assumptions of Theorem 4.6 are satisfied and in addition that f is twice continuously differentiable in the level set S. Suppose that $B_k = \nabla^2 f(x_k)$ for all k, and that the approximate solution p_k of (4.3) at each iteration satisfies (4.52) for some fixed $\gamma > 0$. Then $\lim_{k\to\infty} \|g_k\| = 0$.

If, in addition, the level set S of (4.24) is compact, then either the algorithm terminates at a point x_k at which the second-order necessary conditions (Theorem 2.3) for a local solution hold, or else $\{x_k\}$ has a limit point x^* in S at which the second-order necessary conditions hold.

Fast Local Convergence

Theorem 4.9.

Let f be twice Lipschitz continuously differentiable in a neighborhhod of a point x^* at which second-order sufficient conditions (Theorem 2.4) are satisfied. Suppose the sequence $\{x_k\}$ converges to x^* and that for all k sufficiently large, the trust-region algorithm based on (4.3) with $B_k = \nabla^2 f(x_k)$ chooses steps p_k that satisfy the Cauchy-point-based model reduction criterion (4.20) and are asymptotically similar to Newton steps p_k^{N} whenever $\|p_k^{\mathrm{N}}\| \le \frac{1}{2}\Delta_k$, that is,

$$\|p_k - p_k^{\mathrm{N}}\| = o(\|p_k^{\mathrm{N}}\|). \tag{4.53}$$

Then the trust-region bound Δ_k becomes inactive for all k sufficiently large and the sequence $\{x_k\}$ converges superlinearly to x^*.

How do we solve the subproblem?

- Very sophisticated approach based on theorem on structure of TR solution, eigenvalue analysis and/or an "inner" Newton iteration.
- Foundation: Find Solution for

$$p(\lambda) = -(B + \lambda I)^{-1} g$$

$$\| p(\lambda) \| = \Delta .$$

How do I find such a solution?

$$B = Q\Lambda Q^T \qquad \Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n),$$

$$p(\lambda) = -Q(\Lambda + \lambda I)^{-1} Q^T g = -\sum_{j=1}^{n} \frac{q_j^T g}{\lambda_j + \lambda} q_j,$$

, by orthonormality of $q_1, q_2, \ldots, q_n$:

$$\|p(\lambda)\|^2 = \sum_{j=1}^{n} \frac{\left(q_j^T g\right)^2}{(\lambda_j + \lambda)^2}.$$

CHICAGO

TR problem has a solution

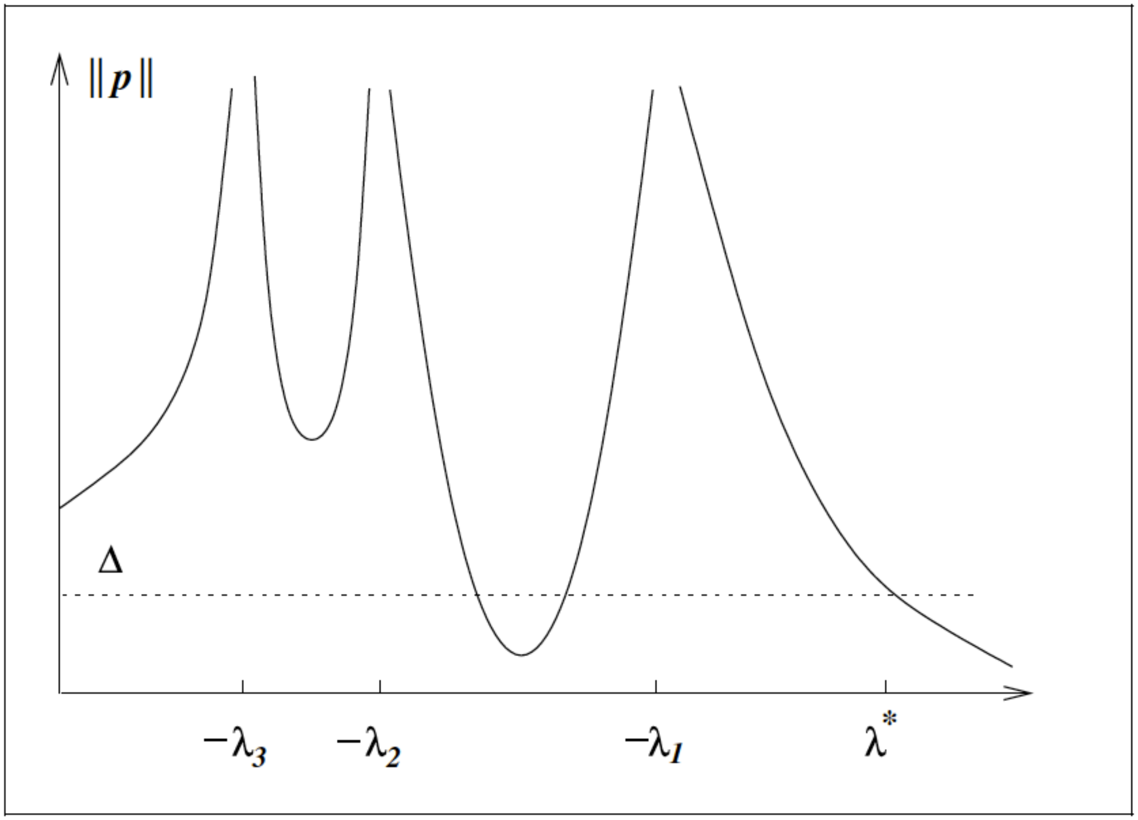


Figure 4.5 $\|p(\lambda)\|$ as a function of λ.

$$\lim_{\lambda \to \infty} \|p(\lambda)\| = 0. \qquad q_j^T g \neq 0 \;\Rightarrow\; \lim_{\lambda \to -\lambda_j} \|p(\lambda)\| = \infty.$$

Practical (INCOMPLETE) algorithm

$$\phi_2(\lambda) = \frac{1}{\Delta} - \frac{1}{\|p(\lambda)\|}, \qquad \lambda^{(\ell+1)} = \lambda^{(\ell)} - \frac{\phi_2\left(\lambda^{(\ell)}\right)}{\phi_2'\left(\lambda^{(\ell)}\right)}.$$

Algorithm 4.3 (Trust Region Subproblem).

Given $\lambda^{(0)}$, $\Delta > 0$:

for $\ell = 0, 1, 2, \ldots$

Factor $B + \lambda^{(\ell)} I = R^T R$;

Solve $R^T R p_\ell = -g$, $R^T q_\ell = p_\ell$;

Set

$$\lambda^{(\ell+1)} = \lambda^{(\ell)} + \left(\frac{\|p_\ell\|}{\|q_\ell\|}\right)^2 \left(\frac{\|p_\ell\| - \Delta}{\Delta}\right);$$

end (for).

It generally gives a machine precision solution in 2-3 iterations (Cholesky)

The Hard Case

$$q_j^T g = 0$$

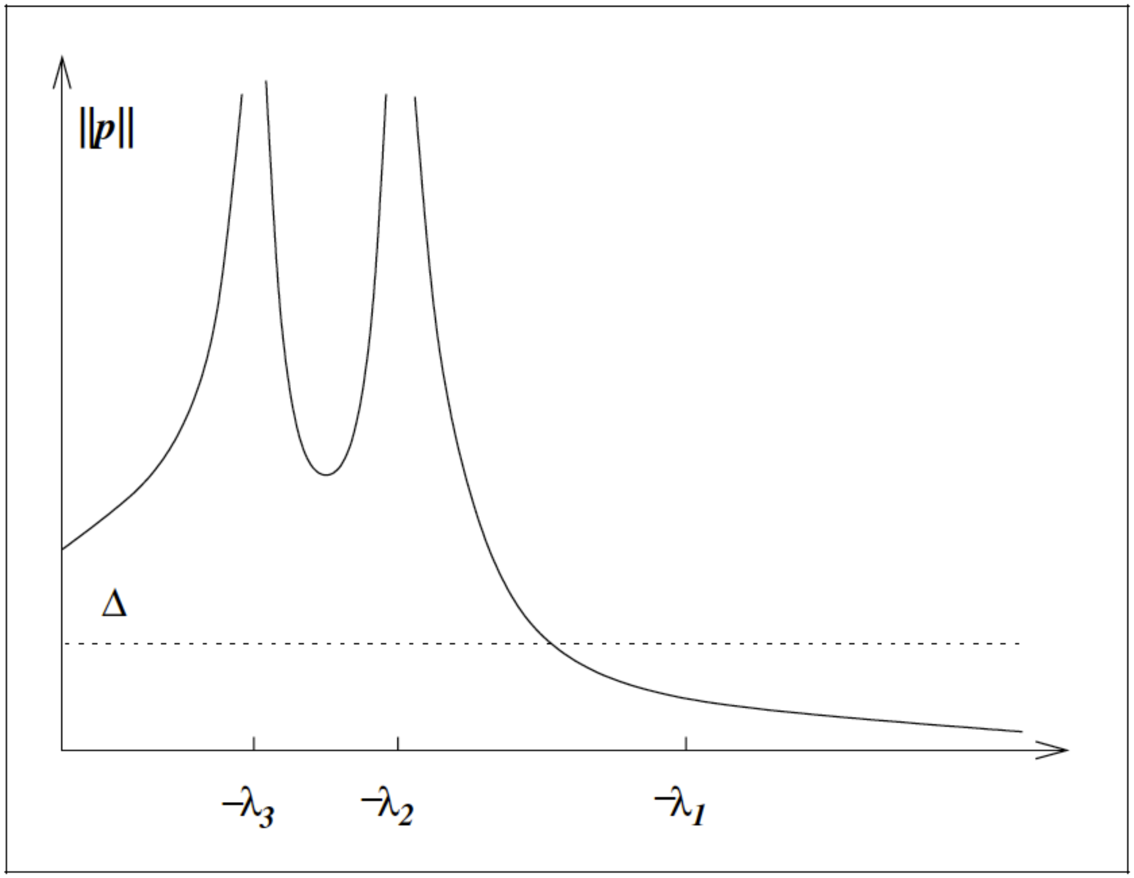


Figure 4.7 The hard case: $\|p(\lambda)\| < \Delta$ for all $\lambda \in (-\lambda_1, \infty)$.

$$\lambda = -\lambda_1 \Rightarrow p = \sum_{j:\lambda_j \neq \lambda_1} \frac{q_j^T g}{\lambda_j - \lambda_1} q_j$$

$$p(\tau) = \sum_{j:\lambda_j \neq \lambda_1} \frac{q_j^T g}{\lambda_j - \lambda_1} q_j + \tau q_1$$

$$\exists \tau \quad \|p(\tau)\| = \Delta^k$$

If double root, things continue to be complicated …

Summary and Comparisons

- Line search problems have easier subproblems (if we modify Cholesky).
- But they cannot be guaranteed to converge to a point with positive semidefinite Hessian.
- Trust-region problems can, at the cost of solving a complicated subproblem.
- Dogleg methods leave "between" these two situations.