

Part 3: Trust-region methods for unconstrained optimization

Nick Gould (RAL)

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x)$$

MSc course on nonlinear optimization

UNCONSTRAINED MINIMIZATION

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x)$$

where the **objective function** $f : \mathbb{R}^n \rightarrow \mathbb{R}$

- ⊙ assume that $f \in C^1$ (sometimes C^2) and Lipschitz
- ⊙ often in practice this assumption violated, but not necessary

LINESEARCH VS TRUST-REGION METHODS

⊙ Linesearch methods

- ◇ pick descent direction p_k
- ◇ pick stepsize α_k to “reduce” $f(x_k + \alpha p_k)$
- ◇ $x_{k+1} = x_k + \alpha_k p_k$

⊙ Trust-region methods

- ◇ pick step s_k to reduce “model” of $f(x_k + s)$
- ◇ accept $x_{k+1} = x_k + s_k$ if decrease in model inherited by $f(x_k + s_k)$
- ◇ otherwise set $x_{k+1} = x_k$, “refine” model

TRUST-REGION MODEL PROBLEM

Model $f(x_k + s)$ by:

- ⊙ linear model

$$m_k^L(s) = f_k + s^T g_k$$

- ⊙ quadratic model — symmetric B_k

$$m_k^Q(s) = f_k + s^T g_k + \frac{1}{2} s^T B_k s$$

Major difficulties:

- ⊙ models may not resemble $f(x_k + s)$ if s is large
- ⊙ models may be unbounded from below
 - ◇ linear model - always unless $g_k = 0$
 - ◇ quadratic model - always if B_k is indefinite, possibly if B_k is only positive semi-definite

THE TRUST REGION

Prevent model $m_k(s)$ from unboundedness by imposing a **trust-region** constraint

$$\|s\| \leq \Delta_k$$

for some “suitable” scalar **radius** $\Delta_k > 0$

\implies **trust-region subproblem**

$$\text{approx minimize } m_k(s) \text{ subject to } \|s\| \leq \Delta_k \\ s \in \mathbb{R}^n$$

- ⊙ in theory does not depend on norm $\|\cdot\|$
- ⊙ in practice it might!

OUR MODEL

For simplicity, concentrate on the second-order (Newton-like) model

$$m_k(s) = m_k^Q(s) = f_k + s^T g_k + \frac{1}{2} s^T B_k s$$

and any trust-region norm $\|\cdot\|$ for which

$$\kappa_s \|\cdot\| \leq \|\cdot\|_2 \leq \kappa_l \|\cdot\|$$

for some $\kappa_l \geq \kappa_s > 0$

Note:

- ⊙ $B_k = H_k$ is allowed
- ⊙ important norms in \mathbb{R}^n
 - ◇ $\|\cdot\|_2 \leq \|\cdot\|_2 \leq \|\cdot\|_2$ (!!!)
 - ◇ $n^{-\frac{1}{2}} \|\cdot\|_1 \leq \|\cdot\|_2 \leq \|\cdot\|_1$
 - ◇ $\|\cdot\|_\infty \leq \|\cdot\|_2 \leq n \|\cdot\|_\infty$

BASIC TRUST-REGION METHOD

Given $k = 0$, $\Delta_0 > 0$ and x_0 , until “convergence” do:

Build the second-order model $m(s)$ of $f(x_k + s)$.

“Solve” the trust-region subproblem to find s_k

for which $m(s_k)$ “<” f_k and $\|s_k\| \leq \Delta_k$, and define

$$\rho_k = \frac{f_k - f(x_k + s_k)}{f_k - m_k(s_k)}.$$

If $\rho_k \geq \eta_v$ [**very successful**]

$$0 < \eta_v < 1$$

set $x_{k+1} = x_k + s_k$ and $\Delta_{k+1} = \gamma_i \Delta_k$

$$\gamma_i \geq 1$$

Otherwise if $\rho_k \geq \eta_s$ then [**successful**]

$$0 < \eta_s \leq \eta_v < 1$$

set $x_{k+1} = x_k + s_k$ and $\Delta_{k+1} = \Delta_k$

Otherwise [**unsuccessful**]

set $x_{k+1} = x_k$ and $\Delta_{k+1} = \gamma_d \Delta_k$

$$0 < \gamma_d < 1$$

Increase k by 1

“SOLVE” THE TRUST REGION SUBPROBLEM?

At the very least

- ⊙ aim to achieve as much reduction in the model as would an iteration of steepest descent
- ⊙ **Cauchy point**: $s_k^c = -\alpha_k^c g_k$ where

$$\alpha_k^c = \arg \min_{\alpha > 0} m_k(-\alpha g_k) \quad \text{subject to} \quad \alpha \|g_k\| \leq \Delta_k$$

$$= \arg \min_{0 < \alpha \leq \Delta_k / \|g_k\|} m_k(-\alpha g_k)$$

◇ minimize quadratic on line segment \implies very easy!

- ⊙ require that

$$m_k(s_k) \leq m_k(s_k^c) \quad \text{and} \quad \|s_k\| \leq \Delta_k$$

- ⊙ in practice, hope to do far better than this

ACHIEVABLE MODEL DECREASE

Theorem 3.1. If $m_k(s)$ is the second-order model and s_k^C is its Cauchy point within the trust-region $\|s\| \leq \Delta_k$,

$$f_k - m_k(s_k^C) \geq \frac{1}{2}\|g_k\|_2 \min \left[\frac{\|g_k\|_2}{1 + \|B_k\|_2}, \kappa_s \Delta_k \right].$$

PROOF OF THEOREM 3.1

$$m_k(-\alpha g_k) = f_k - \alpha \|g_k\|_2^2 + \frac{1}{2}\alpha^2 g_k^T B_k g_k.$$

Result immediate if $g_k = 0$.

Otherwise, 3 possibilities

- (i) curvature $g_k^T B_k g_k \leq 0 \implies m_k(-\alpha g_k)$ unbounded from below as α increases \implies Cauchy point occurs on the trust-region boundary.
- (ii) curvature $g_k^T B_k g_k > 0$ & minimizer $m_k(-\alpha g_k)$ occurs at or beyond the trust-region boundary \implies Cauchy point occurs on the trust-region boundary.
- (iii) the curvature $g_k^T B_k g_k > 0$ & minimizer $m_k(-\alpha g_k)$, and hence Cauchy point, occurs before trust-region is reached.

Consider each case in turn;

Case (i)

$$g_k^T B_k g_k \leq 0 \ \& \ \alpha \geq 0 \implies$$

$$m_k(-\alpha g_k) = f_k - \alpha \|g_k\|_2^2 + \frac{1}{2} \alpha^2 g_k^T B_k g_k \leq f_k - \alpha \|g_k\|_2^2 \quad (1)$$

Cauchy point lies on boundary of the trust region \implies

$$\alpha_k^C = \frac{\Delta_k}{\|g_k\|}. \quad (2)$$

$$(1) + (2) \implies$$

$$f_k - m_k(s_k^C) \geq \|g_k\|_2^2 \frac{\Delta_k}{\|g_k\|} \geq \kappa_s \|g_k\|_2 \Delta_k \geq \frac{1}{2} \kappa_s \|g_k\|_2 \Delta_k$$

since $\|g_k\|_2 \geq \kappa_s \|g_k\|$.

Case (ii)

$$\alpha_k^* \stackrel{\text{def}}{=} \arg \min m_k(-\alpha g_k) \equiv f_k - \alpha \|g_k\|_2^2 + \frac{1}{2} \alpha^2 g_k^T B_k g_k \quad (3)$$

\implies

$$\alpha_k^* = \frac{\|g_k\|_2^2}{g_k^T B_k g_k} \geq \alpha_k^C = \frac{\Delta_k}{\|g_k\|} \quad (4)$$

\implies

$$\alpha_k^C g_k^T B_k g_k \leq \|g_k\|_2^2. \quad (5)$$

$$(3) + (4) + (5) \ \& \ \|g_k\|_2 \geq \kappa_s \|g_k\| \implies$$

$$\begin{aligned} f_k - m_k(s_k^C) &= \alpha_k^C \|g_k\|_2^2 - \frac{1}{2} [\alpha_k^C]^2 g_k^T B_k g_k \geq \frac{1}{2} \alpha_k^C \|g_k\|_2^2 \\ &= \frac{1}{2} \|g_k\|_2^2 \frac{\Delta_k}{\|g_k\|} \geq \frac{1}{2} \kappa_s \|g_k\|_2 \Delta_k. \end{aligned}$$

Case (iii)

$$\alpha_k^c = \alpha_k^* = \frac{\|g_k\|_2^2}{g_k^T B_k g_k}$$

\implies

$$\begin{aligned} f_k - m_k(s_k^c) &= \alpha_k^* \|g_k\|_2^2 + \frac{1}{2} (\alpha_k^*)^2 g_k^T B_k g_k \\ &= \frac{\|g_k\|_2^4}{g_k^T B_k g_k} - \frac{1}{2} \frac{\|g_k\|_2^4}{g_k^T B_k g_k} \\ &= \frac{1}{2} \frac{\|g_k\|_2^4}{g_k^T B_k g_k} \\ &\geq \frac{1}{2} \frac{\|g_k\|_2^2}{1 + \|B_k\|_2}, \end{aligned}$$

where

$$|g_k^T B_k g_k| \leq \|g_k\|_2^2 \|B_k\|_2 \leq \|g_k\|_2^2 (1 + \|B_k\|_2)$$

because of the Cauchy-Schwarz inequality.

Corollary 3.2. If $m_k(s)$ is the second-order model, and s_k is an improvement on the Cauchy point within the trust-region $\|s\| \leq \Delta_k$,

$$f_k - m_k(s_k) \geq \frac{1}{2} \|g_k\|_2 \min \left[\frac{\|g_k\|_2}{1 + \|B_k\|_2}, \kappa_s \Delta_k \right].$$

DIFFERENCE BETWEEN MODEL AND FUNCTION

Lemma 3.3. Suppose that $f \in C^2$, and that the true and model Hessians satisfy the bounds $\|H(x)\|_2 \leq \kappa_h$ for all x and $\|B_k\|_2 \leq \kappa_b$ for all k and some $\kappa_h \geq 1$ and $\kappa_b \geq 0$. Then

$$|f(x_k + s_k) - m_k(s_k)| \leq \kappa_d \Delta_k^2,$$

where $\kappa_d = \frac{1}{2}\kappa_l^2(\kappa_h + \kappa_b)$, for all k .

PROOF OF LEMMA 3.3

Mean value theorem \implies

$$f(x_k + s_k) = f(x_k) + s_k^T \nabla_x f(x_k) + \frac{1}{2} s_k^T \nabla_{xx} f(\xi_k) s_k$$

for some $\xi_k \in [x_k, x_k + s_k]$. Thus

$$\begin{aligned} |f(x_k + s_k) - m_k(s_k)| &= \left| \frac{1}{2} s_k^T H(\xi_k) s_k - s_k^T B_k s_k \right| \leq \frac{1}{2} |s_k^T H(\xi_k) s_k| + \frac{1}{2} |s_k^T B_k s_k| \\ &\leq \frac{1}{2} (\kappa_h + \kappa_b) \|s_k\|_2^2 \leq \frac{1}{2} \kappa_l^2 (\kappa_h + \kappa_b) \|s_k\|_2^2 \leq \kappa_d \Delta_k^2 \end{aligned}$$

using the triangle and Cauchy-Schwarz inequalities.

ULTIMATE PROGRESS AT NON-OPTIMAL POINTS

Lemma 3.4. Suppose that $f \in C^2$, that the true and model Hessians satisfy the bounds $\|H_k\|_2 \leq \kappa_h$ and $\|B_k\|_2 \leq \kappa_b$ for all k and some $\kappa_h \geq 1$ and $\kappa_b \geq 0$, and that $\kappa_d = \frac{1}{2}\kappa_l^2(\kappa_h + \kappa_b)$. Suppose furthermore that $g_k \neq 0$ and that

$$\Delta_k \leq \|g_k\|_2 \min\left(\frac{1}{\kappa_s(\kappa_h + \kappa_b)}, \frac{\kappa_s(1 - \eta_v)}{2\kappa_d}\right).$$

Then iteration k is very successful and

$$\Delta_{k+1} \geq \Delta_k.$$

PROOF OF LEMMA 3.4

By definition,

$$1 + \|B_k\|_2 \leq \kappa_h + \kappa_b$$

+ first bound on $\Delta_k \implies$

$$\kappa_s \Delta_k \leq \frac{\|g_k\|_2}{\kappa_h + \kappa_b} \leq \frac{\|g_k\|_2}{1 + \|B_k\|_2}.$$

Corollary 3.2 \implies

$$f_k - m_k(s_k) \geq \frac{1}{2}\|g_k\|_2 \min\left[\frac{\|g_k\|_2}{1 + \|B_k\|_2}, \kappa_s \Delta_k\right] = \frac{1}{2}\kappa_s \|g_k\|_2 \Delta_k.$$

+ Lemma 3.3 + second bound on $\Delta_k \implies$

$$|\rho_k - 1| = \left| \frac{f(x_k + s_k) - m_k(s_k)}{f_k - m_k(s_k)} \right| \leq 2 \frac{\kappa_d \Delta_k^2}{\kappa_s \|g_k\|_2 \Delta_k} = \frac{2\kappa_d}{\kappa_s} \frac{\Delta_k}{\|g_k\|_2} \leq 1 - \eta_v.$$

$\implies \rho_k \geq \eta_v \implies$ iteration is very successful.

RADIUS WON'T SHRINK TO ZERO AT NON-OPTIMAL POINTS

Lemma 3.5. Suppose that $f \in C^2$, that the true and model Hessians satisfy the bounds $\|H_k\|_2 \leq \kappa_h$ and $\|B_k\|_2 \leq \kappa_b$ for all k and some $\kappa_h \geq 1$ and $\kappa_b \geq 0$, and that $\kappa_d = \frac{1}{2}\kappa_l^2(\kappa_h + \kappa_b)$. Suppose furthermore that there exists a constant $\epsilon > 0$ such that $\|g_k\|_2 \geq \epsilon$ for all k . Then

$$\Delta_k \geq \kappa_\epsilon \stackrel{\text{def}}{=} \epsilon \gamma_d \min \left(\frac{1}{\kappa_s(\kappa_h + \kappa_b)}, \frac{\kappa_s(1 - \eta_v)}{2\kappa_d} \right)$$

for all k .

PROOF OF LEMMA 3.5

Suppose otherwise that iteration k is first for which

$$\Delta_{k+1} \leq \kappa_\epsilon.$$

$\Delta_k > \Delta_{k+1} \implies$ iteration k unsuccessful $\implies \gamma_d \Delta_k \leq \Delta_{k+1}$. Hence

$$\begin{aligned} \Delta_k &\leq \epsilon \min \left(\frac{1}{\kappa_s(\kappa_h + \kappa_b)}, \frac{\kappa_s(1 - \eta_v)}{2\kappa_d} \right) \\ &\leq \|g_k\|_2 \min \left(\frac{1}{\kappa_s(\kappa_h + \kappa_b)}, \frac{\kappa_s(1 - \eta_v)}{2\kappa_d} \right) \end{aligned}$$

But this contradicts assertion of Lemma 3.4 that iteration k must be very successful.

POSSIBLE FINITE TERMINATION

Lemma 3.6. Suppose that $f \in C^2$, and that both the true and model Hessians remain bounded for all k . Suppose furthermore that there are only finitely many successful iterations. Then $x_k = x_*$ for all sufficiently large k and $g(x_*) = 0$.

PROOF OF LEMMA 3.6

$$x_{k_0+j} = x_{k_0+1} = x_*$$

for all $j > 0$, where k_0 is index of last successful iterate.

All iterations are unsuccessful for sufficiently large $k \implies \{\Delta_k\} \longrightarrow 0$

+ Lemma 3.4 then implies that if $\|g_{k_0+1}\| > 0$ there must be a successful iteration of index larger than k_0 , which is impossible $\implies \|g_{k_0+1}\| = 0$.

GLOBAL CONVERGENCE OF ONE SEQUENCE

Theorem 3.7. Suppose that $f \in C^2$, and that both the true and model Hessians remain bounded for all k . Then either

$$g_l = 0 \text{ for some } l \geq 0$$

or

$$\lim_{k \rightarrow \infty} f_k = -\infty$$

or

$$\liminf_{k \rightarrow \infty} \|g_k\| = 0.$$

PROOF OF THEOREM 3.7

Let \mathcal{S} be the index set of successful iterations. Lemma 3.6 \implies true Theorem 3.7 when $|\mathcal{S}|$ finite.

So consider $|\mathcal{S}| = \infty$, and suppose f_k bounded below and

$$\|g_k\|_2 \geq \epsilon \tag{6}$$

for some $\epsilon > 0$ and all k , and consider some $k \in \mathcal{S}$.

+ Corollary 3.2, Lemma 3.5, and the assumption (6) \implies

$$f_k - f_{k+1} \geq \eta_s [f_k - m_k(s_k)] \geq \delta_\epsilon \stackrel{\text{def}}{=} \frac{1}{2} \eta_s \epsilon \min \left[\frac{\epsilon}{1 + \kappa_b}, \kappa_s \kappa_\epsilon \right].$$

\implies

$$f_0 - f_{k+1} = \sum_{\substack{j=0 \\ j \in \mathcal{S}}}^k [f_j - f_{j+1}] \geq \sigma_k \delta_\epsilon,$$

where σ_k is the number of successful iterations up to iteration k . But

$$\lim_{k \rightarrow \infty} \sigma_k = +\infty.$$

$\implies f_k$ unbounded below \implies a subsequence of the $\|g_k\|_2 \longrightarrow 0$

GLOBAL CONVERGENCE

Theorem 3.8. Suppose that $f \in C^2$, and that both the true and model Hessians remain bounded for all k . Then either

$$g_l = 0 \text{ for some } l \geq 0$$

or

$$\lim_{k \rightarrow \infty} f_k = -\infty$$

or

$$\lim_{k \rightarrow \infty} g_k = 0.$$

PROOF OF THEOREM 3.8

Suppose otherwise that f_k is bounded from below, and that there is a subsequence $\{t_i\} \subseteq \mathcal{S}$, such that

$$\|g_{t_i}\|_2 \geq 2\epsilon > 0 \tag{7}$$

for some $\epsilon > 0$ and for all i . Theorem 3.7 $\implies \exists \{\ell_i\} \subseteq \mathcal{S}$ such that

$$\|g_k\| \geq \epsilon \text{ for } t_i \leq k < \ell_i \text{ and } \|g_{\ell_i}\|_2 < \epsilon. \tag{8}$$

Now restrict attention to indices in

$$\mathcal{K} \stackrel{\text{def}}{=} \{k \in \mathcal{S} \mid t_i \leq k < \ell_i\}.$$

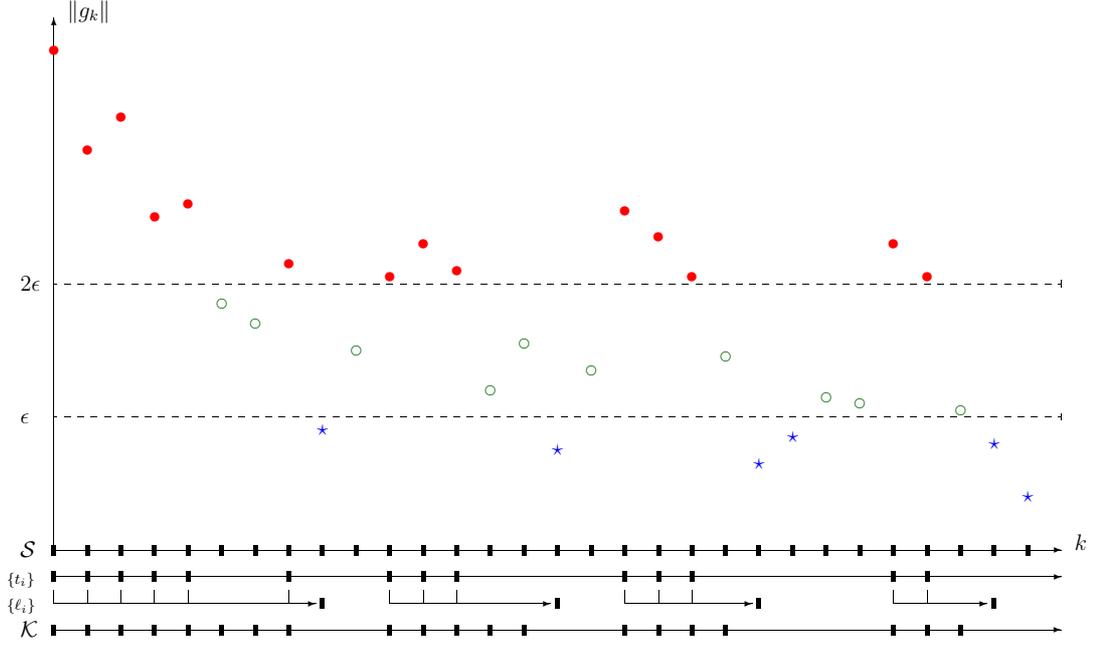


Figure 3.1: The subsequences of the proof of Theorem 3.8

As in proof of Theorem 3.7, (8) \implies

$$f_k - f_{k+1} \geq \eta_s [f_k - m_k(s_k)] \geq \frac{1}{2} \eta_s \epsilon \min \left[\frac{\epsilon}{1 + \kappa_b}, \kappa_s \Delta_k \right] \quad (9)$$

for all $k \in \mathcal{K} \implies$ LHS of (9) $\longrightarrow 0$ as $k \longrightarrow \infty \implies$

$$\lim_{\substack{k \rightarrow \infty \\ k \in \mathcal{K}}} \Delta_k = 0$$

\implies

$$\Delta_k \leq \frac{2}{\epsilon \eta_s \kappa_s} [f_k - f_{k+1}].$$

for $k \in \mathcal{K}$ sufficiently large \implies

$$\|x_{t_i} - x_{l_i}\| \leq \sum_{\substack{j=t_i \\ j \in \mathcal{K}}}^{\ell_i-1} \|x_j - x_{j+1}\| \leq \sum_{\substack{j=t_i \\ j \in \mathcal{K}}}^{\ell_i-1} \Delta_j \leq \frac{2}{\epsilon \eta_s \kappa_s} [f_{t_i} - f_{l_i}]. \quad (10)$$

for i sufficiently large.

But RHS of (10) $\longrightarrow 0 \implies \|x_{t_i} - x_{l_i}\| \longrightarrow 0$ as i tends to infinity
+ continuity $\implies \|g_{t_i} - g_{l_i}\| \longrightarrow 0$.

Impossible as $\|g_{t_i} - g_{\ell_i}\| \geq \epsilon$ by definition of $\{t_i\}$ and $\{\ell_i\} \implies$ no subsequence satisfying (7) can exist.

II: SOLVING THE TRUST-REGION SUBPROBLEM

(approximately) minimize $q(s) \equiv s^T g + \frac{1}{2} s^T B s$ subject to $\|s\| \leq \Delta$
 $s \in \mathbb{R}^n$

AIM: find s_* so that

$$q(s_*) \leq q(s^c) \text{ and } \|s_*\| \leq \Delta$$

Might solve

- ⊙ exactly \implies Newton-like method
- ⊙ approximately \implies steepest descent/conjugate gradients

THE ℓ_2 -NORM TRUST-REGION SUBPROBLEM

$$\underset{s \in \mathbb{R}^n}{\text{minimize}} \quad q(s) \equiv s^T g + \frac{1}{2} s^T B s \quad \text{subject to} \quad \|s\|_2 \leq \Delta$$

Solution characterisation result:

Theorem 3.9. Any *global* minimizer s_* of $q(s)$ subject to $\|s\|_2 \leq \Delta$ satisfies the equation

$$(B + \lambda_* I) s_* = -g,$$

where $B + \lambda_* I$ is positive semi-definite, $\lambda_* \geq 0$ and $\lambda_*(\|s_*\|_2 - \Delta) = 0$. If $B + \lambda_* I$ is positive definite, s_* is unique.

PROOF OF THEOREM 3.9

Problem equivalent to minimizing $q(s)$ subject to $\frac{1}{2}\Delta^2 - \frac{1}{2}s^T s \geq 0$.
Theorem 1.9 \implies

$$g + B s_* = -\lambda_* s_* \tag{11}$$

for some Lagrange multiplier $\lambda_* \geq 0$ for which either $\lambda_* = 0$ or $\|s_*\|_2 = \Delta$ (or both). It remains to show $B + \lambda_* I$ is positive semi-definite.

If s_* lies in the interior of the trust-region, $\lambda_* = 0$, and Theorem 1.10 $\implies B + \lambda_* I = B$ is positive semi-definite.

If $\|s_*\|_2 = \Delta$ and $\lambda_* = 0$, Theorem 1.10 $\implies v^T B v \geq 0$ for all $v \in \mathcal{N}_+ = \{v | s_*^T v \geq 0\}$. If $v \notin \mathcal{N}_+ \implies -v \in \mathcal{N}_+ \implies v^T B v \geq 0$ for all v .

Only remaining case is where $\|s_*\|_2 = \Delta$ and $\lambda_* > 0$. Theorem 1.10 $\implies v^T (B + \lambda_* I) v \geq 0$ for all $v \in \mathcal{N}_+ = \{v | s_*^T v = 0\} \implies$ remains to consider $v^T B v$ when $s_*^T v \neq 0$.

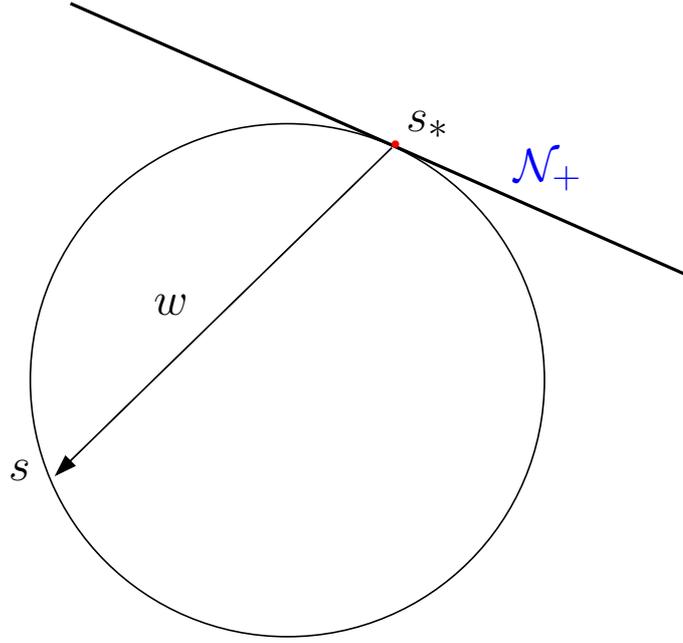


Figure 3.2: Construction of “missing” directions of positive curvature.

Let s be any point on the boundary δR of the trust-region R , and let $w = s - s_*$. Then

$$-w^T s_* = (s_* - s)^T s_* = \frac{1}{2}(s_* - s)^T (s_* - s) = \frac{1}{2}w^T w \quad (12)$$

since $\|s\|_2 = \Delta = \|s_*\|_2$. (11) + (12) \implies

$$\begin{aligned} q(s) - q(s_*) &= w^T (g + B s_*) + \frac{1}{2}w^T B w \\ &= -\lambda_* w^T s_* + \frac{1}{2}w^T B w = \frac{1}{2}w^T (B + \lambda_* I) w, \end{aligned} \quad (13)$$

$\implies w^T (B + \lambda_* I) w \geq 0$ since s_* is a global minimizer. But

$$s = s_* - 2 \frac{s_*^T v}{v^T v} v \in \delta R$$

\implies (for this s) $w \parallel v \implies v^T (B + \lambda_* I) v \geq 0$.

When $B + \lambda_* I$ is positive definite, $s_* = -(B + \lambda_* I)^{-1} g$. If $s_* \in \delta R$ and $s \in R$, (12) and (13) become $-w^T s_* \geq \frac{1}{2}w^T w$ and $q(s) \geq q(s_*) + \frac{1}{2}w^T (B + \lambda_* I) w$ respectively. Hence, $q(s) > q(s_*)$ for any $s \neq s_*$. If s_* is interior, $\lambda_* = 0$, B is positive definite, and thus s_* is the unique unconstrained minimizer of $q(s)$.

ALGORITHMS FOR THE ℓ_2 -NORM SUBPROBLEM

Two cases:

◦ B positive-semi definite and $Bs = -g$ satisfies $\|s\|_2 \leq \Delta \implies s_* = s$

◦ B indefinite or $Bs = -g$ satisfies $\|s\|_2 > \Delta$

In this case

◊ $(B + \lambda_* I)s_* = -g$ and $s_*^T s_* = \Delta^2$

◊ nonlinear (quadratic) system in s and λ

◊ concentrate on this

EQUALITY CONSTRAINED ℓ_2 -NORM SUBPROBLEM

Suppose B has spectral decomposition

$$B = U^T \Lambda U$$

◦ U eigenvectors

◦ Λ diagonal eigenvalues: $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$

Require $B + \lambda I$ positive semi-definite $\implies \lambda \geq -\lambda_1$

Define

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

Require

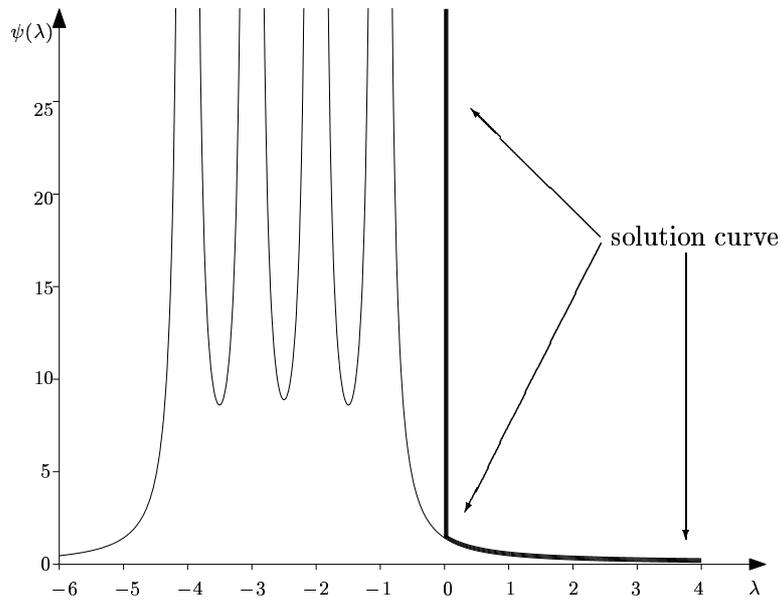
$$\psi(\lambda) \stackrel{\text{def}}{=} \|s(\lambda)\|_2^2 = \Delta^2$$

Note

$$(\gamma_i = e_i^T U g)$$

$$\psi(\lambda) = \|U^T(\Lambda + \lambda I)^{-1}Ug\|_2^2 = \sum_{i=1}^n \frac{\gamma_i^2}{(\lambda_i + \lambda)^2}$$

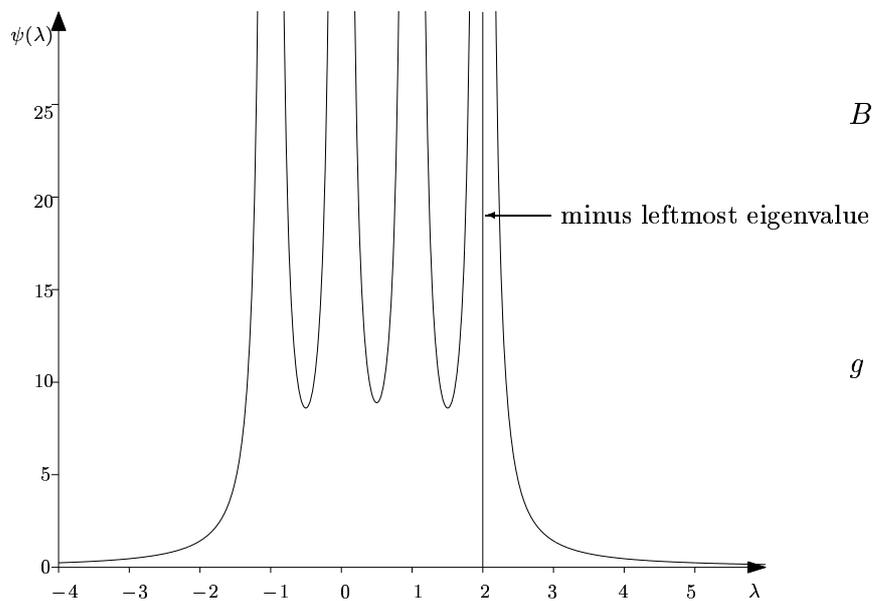
CONVEX EXAMPLE



$$B = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix}$$

$$g = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

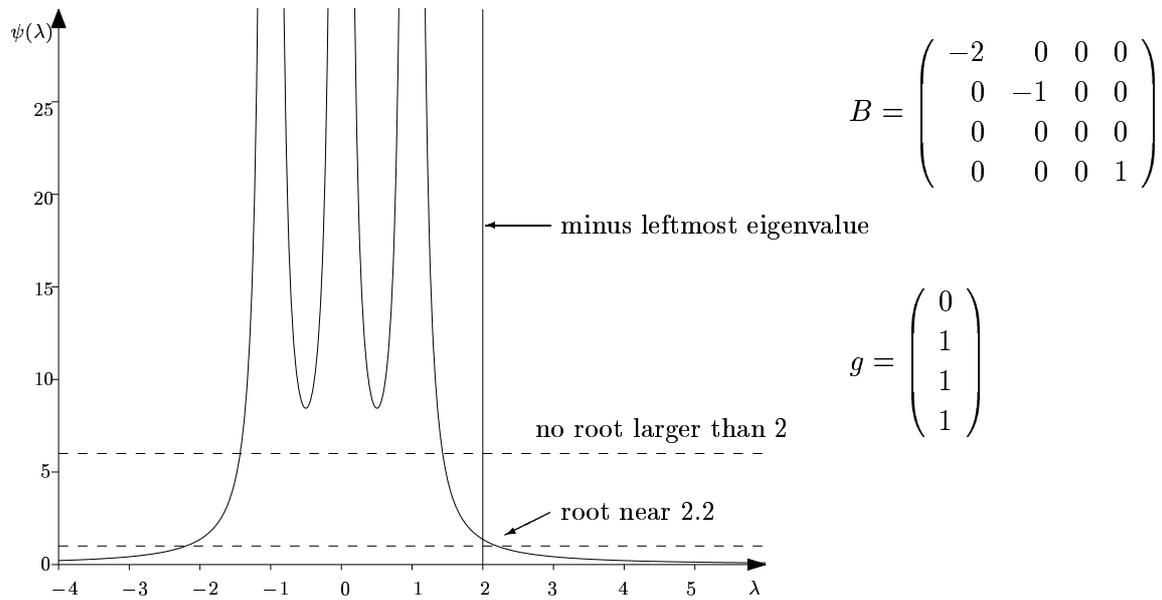
NONCONVEX EXAMPLE



$$B = \begin{pmatrix} -2 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$g = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

THE “HARD” CASE



SUMMARY

For indefinite B ,

Hard case occurs when g orthogonal to eigenvector u_1
 for most negative eigenvalue λ_1

- ⊙ OK if radius is radius small enough
- ⊙ No “obvious” solution to equations ... but
 solution is actually of the form

$$s_{\text{lim}} + \sigma u_1$$

where

- ◇ $s_{\text{lim}} = \lim_{\lambda \rightarrow -\lambda_1} s(\lambda)$
- ◇ $\|s_{\text{lim}} + \sigma u_1\|_2 = \Delta$

HOW TO SOLVE $\|s(\lambda)\|_2 = \Delta$

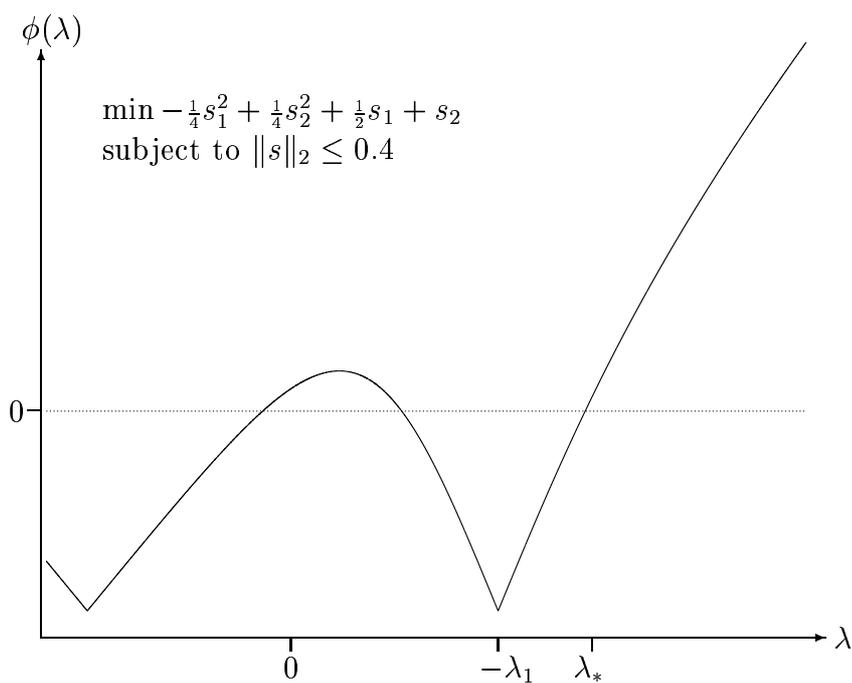
DON'T!!

Solve instead the **secular equation**

$$\phi(\lambda) \stackrel{\text{def}}{=} \frac{1}{\|s(\lambda)\|_2} - \frac{1}{\Delta} = 0$$

- ⊙ no poles
- ⊙ smallest at eigenvalues (except in hard case!)
- ⊙ analytic function \implies ideal for Newton
- ⊙ global convergent (ultimately quadratic rate except in hard case)
- ⊙ need to safeguard to protect Newton from the hard & interior solution cases

THE SECULAR EQUATION



NEWTON'S METHOD FOR SECULAR EQUATION

Newton correction at λ is $-\phi(\lambda)/\phi'(\lambda)$. Differentiating

$$\phi(\lambda) = \frac{1}{\|s(\lambda)\|_2} - \frac{1}{\Delta} = \frac{1}{(s^T(\lambda)s(\lambda))^{\frac{1}{2}}} - \frac{1}{\Delta} \implies$$
$$\phi'(\lambda) = -\frac{s^T(\lambda)\nabla_\lambda s(\lambda)}{(s^T(\lambda)s(\lambda))^{\frac{3}{2}}} = -\frac{s^T(\lambda)\nabla_\lambda s(\lambda)}{\|s(\lambda)\|_2^3}.$$

Differentiating the defining equation

$$(B + \lambda I)s(\lambda) = -g \implies (B + \lambda I)\nabla_\lambda s(\lambda) + s(\lambda) = 0.$$

Notice that, rather than $\nabla_\lambda s(\lambda)$, merely

$$s^T(\lambda)\nabla_\lambda s(\lambda) = -s^T(\lambda)(B + \lambda I)^{-1}s(\lambda)$$

required for $\phi'(\lambda)$. Given the factorization $B + \lambda I = L(\lambda)L^T(\lambda) \implies$

$$s^T(\lambda)(B + \lambda I)^{-1}s(\lambda) = s^T(\lambda)L^{-T}(\lambda)L^{-1}(\lambda)s(\lambda)$$
$$= (L^{-1}(\lambda)s(\lambda))^T(L^{-1}(\lambda)s(\lambda)) = \|w(\lambda)\|_2^2$$

where $L(\lambda)w(\lambda) = s(\lambda)$.

NEWTON'S METHOD & THE SECULAR EQUATION

Let $\lambda > -\lambda_1$ and $\Delta > 0$ be given

Until "convergence" do:

Factorize $B + \lambda I = LL^T$

Solve $LL^T s = -g$

Solve $Lw = s$

Replace λ by

$$\lambda + \left(\frac{\|s\|_2 - \Delta}{\Delta} \right) \left(\frac{\|s\|_2^2}{\|w\|_2^2} \right)$$

SOLVING THE LARGE-SCALE PROBLEM

- ⊙ when n is large, factorization may be impossible
- ⊙ may instead try to use an iterative method to approximate
 - ◇ Steepest descent leads to the Cauchy point
 - ◇ obvious generalization: conjugate gradients ... but
 - ▷ what about the trust region?
 - ▷ what about negative curvature?

CONJUGATE GRADIENTS TO “MINIMIZE” $q(s)$

Given $s^0 = 0$, set $g^0 = g$, $d^0 = -g$ and $i = 0$
Until g^i “small” or breakdown, iterate

$$\alpha^i = \|g^i\|_2^2 / d^{i T} B d^i$$
$$s^{i+1} = s^i + \alpha^i d^i$$
$$g^{i+1} = g^i + \alpha^i B d^i$$
$$\beta^i = \|g^{i+1}\|_2^2 / \|g^i\|_2^2$$
$$d^{i+1} = -g^{i+1} + \beta^i d^i$$

and increase i by 1

Important features

- ⊙ $g^j = B s^j + g$ for all $j = 0, \dots, i$
- ⊙ $d^j T g^{i+1} = 0$ for all $j = 0, \dots, i$
- ⊙ $g^j T g^{i+1} = 0$ for all $j = 0, \dots, i$

CRUCIAL PROPERTY OF CONJUGATE GRADIENTS

Theorem 3.10. Suppose that the conjugate gradient method is applied to minimize $q(s)$ starting from $s^0 = 0$, and that $d^i T B d^i > 0$ for $0 \leq i \leq k$. Then the iterates s^j satisfy the inequalities

$$\|s^j\|_2 < \|s^{j+1}\|_2$$

for $0 \leq j \leq k - 1$.

PROOF OF THEOREM 3.10

First show that

$$d^i T d^j = \frac{\|g^i\|_2^2}{\|g^j\|_2^2} \|d^j\|_2^2 > 0 \quad (14)$$

for all $0 \leq j \leq i \leq k$. For any i , (14) is trivially true for $j = i$. Suppose it is also true for all $i \leq l$. Then, the update for d^{l+1} gives

$$d^{l+1} = -g^{l+1} + \frac{\|g^{l+1}\|_2^2}{\|g^l\|_2^2} d^l.$$

Forming the inner product with d^j , and using the fact that $d^j T g^{l+1} = 0$ for all $j = 0, \dots, l$, and (14) when $j = l$, reveals

$$\begin{aligned} d^{l+1 T} d^j &= -g^{l+1 T} d^j + \frac{\|g^{l+1}\|_2^2}{\|g^l\|_2^2} d^l T d^j \\ &= \frac{\|g^{l+1}\|_2^2 \|g^l\|_2^2}{\|g^l\|_2^2 \|g^j\|_2^2} \|d^j\|_2^2 = \frac{\|g^{l+1}\|_2^2}{\|g^j\|_2^2} \|d^j\|_2^2 > 0. \end{aligned}$$

Thus (14) is true for $i \leq l + 1$, and hence for all $0 \leq j \leq i \leq k$.

Now have from the algorithm that

$$s^i = s^0 + \sum_{j=0}^{i-1} \alpha^j d^j = \sum_{j=0}^{i-1} \alpha^j d^j$$

as, by assumption, $s^0 = 0$. Hence

$$s^{i T} d^i = \sum_{j=0}^{i-1} \alpha^j d^j T d^i = \sum_{j=0}^{i-1} \alpha^j d^j T d^i > 0 \quad (15)$$

as each $\alpha^j > 0$, which follows from the definition of α^j , since $d^j T H d^j > 0$, and from relationship (14). Hence

$$\begin{aligned} \|s^{i+1}\|_2^2 &= s^{i+1 T} s^{i+1} = (s^i + \alpha^i d^i)^T (s^i + \alpha^i d^i) \\ &= s^i T s^i + 2\alpha^i s^i T d^i + \alpha^{i 2} d^i T d^i > s^i T s^i = \|s^i\|_2^2 \end{aligned}$$

follows directly from (15) and $\alpha^i > 0$ which is the required result.

TRUNCATED CONJUGATE GRADIENTS

Apply the conjugate gradient method, but terminate at iteration i if

1. $d^i T B d^i \leq 0 \implies$ problem unbounded along d^i
2. $\|s^i + \alpha^i d^i\|_2 > \Delta \implies$ solution on trust-region boundary

In both cases, stop with $s_* = s^i + \alpha^B d^i$, where α^B chosen as positive root of

$$\|s^i + \alpha^B d^i\|_2 = \Delta$$

Crucially

$$q(s_*) \leq q(s^c) \quad \text{and} \quad \|s_*\|_2 \leq \Delta$$

\implies TR algorithm converges to a first-order critical point

HOW GOOD IS TRUNCATED C.G.?

In the convex case ... very good

Theorem 3.11. Suppose that the truncated conjugate gradient method is applied to minimize $q(s)$ and that B is positive definite. Then the computed and actual solutions to the problem, s_* and s_*^M , satisfy the bound

$$q(s_*) \leq \frac{1}{2}q(s_*^M)$$

In the non-convex case ... maybe poor

- ⊙ e.g., if $g = 0$ and B is indefinite $\implies q(s_*) = 0$

WHAT CAN WE DO IN THE NON-CONVEX CASE?

Solve the problem over a subspace

- ⊙ instead of the B -conjugate subspace for CG, use the equivalent Lanczos orthogonal basis
 - ◇ Gram-Schmidt applied to CG (Krylov) basis \mathcal{D}^i
 - ◇ Subspace $\mathcal{Q}^i = \{s \mid s = Q^i s_q \text{ for some } s_q \in \mathbb{R}^i\}$
 - ◇ Q^i is such that

$$Q^{iT}Q^i = I \text{ and } Q^{iT}BQ^i = T^i$$

where T^i is tridiagonal and $Q^{iT}g = \|g\|_2 e_1$

- ◇ Q^i trivial to generate from CG \mathcal{D}^i

GENERALIZED LANCZOS TRUST-REGION METHOD

$$s^i = \arg \min_{s \in \mathcal{Q}^i} q(s) \text{ subject to } \|s\|_2 \leq \Delta$$

$\implies s^i = Q^i s_q^i$, where

$$s_q^i = \arg \min_{s_q \in \mathbb{R}^i} \|g\|_2 e_1^T s_q + \frac{1}{2} s_q^T T^i s_q \text{ subject to } \|s_q\|_2 \leq \Delta$$

- ⊙ advantage T^i has very sparse factors \implies can solve the problem using the earlier secular equation approach
- ⊙ can exploit all the structure here \implies use solution for one problem to initialize next
- ⊙ until the trust-region boundary is reached, it **is** conjugate gradients \implies switch when we get there