Trust-Region Methods for Smooth Unconstrained Optimization

Daniel P. Robinson
Department of Applied Mathematics and Statistics
Johns Hopkins University

October 6, 2020

Trust region is good for quadratic approximation to non-convex function

constrained optimization { figure out right constrain; reduce to unconstrained using Lagranch

The problem

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

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

- assume that $f \in \mathcal{C}^1$ (sometimes $\mathcal{C}^2$) and is Lipschitz continuous
- in practice this assumption may be violated, but the algorithms we develop may still work
- in practice it is very rare to be able to provide an explicit minimizer
- we consider iterative methods: given starting guess x_0, generate sequence

$$\{x_k\} \quad \text{for } k = 1, 2, \ldots$$

- AIM: ensure that a subsequence has some favorable limiting properties
  - satisfies first-order necessary conditions
  - satisfies second-order necessary conditions

Notation: $f_k = f(x_k)$, $g_k = g(x_k)$, $H_k = H(x_k)$

Outline

Linesearch versus Trust-Region

Linesearch

- considered descent methods, i.e., $f(x_{k+1}) < f(x_k)$
- direction first (search direction p_k), length second (linesearch α_k)
- ensure that this direction is a descent direction, i.e.,

$$g_k^T p_k < 0$$

  so that, for small steps along p_k, the objective function f **will** be reduced
- the computation of α_k may itself require an iterative procedure
- generic update for linesearch methods is given by

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

Trust-region

- will consider descent methods, i.e., $f(x_{k+1}) \leq f(x_k)$
- length first (trust-region radius δ_k), direction second ("solve" subproblem for s_k)
- pick step s_k to reduce some model of $f(x_k + s)$
- generic update is

$$x_{k+1} \leftarrow \begin{cases} x_k + s_k & \text{if } f(x_k + s_k) \text{ "<" } f(x_k) \\ x_k & \text{otherwise} \end{cases}$$

- s_k is used if the decrease in the model is realized by the objective $f(x_k + s_k)$

Models of $f(x_k + s)$

- linear model

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

- quadratic model

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

for some symmetric matrix B_k

Difficulties

- models may not resemble $f(x_k + s)$ when s is large
- ideally, want to choose most accurate second-order model, i.e., $B_k = H_k$
- minimizing the models may not be possible
  - linear model
    - m_k^L is unbounded below unless $g_k = 0$ (already at first-order solution)
  - quadratic model
    - m_k^Q is unbounded below if B_k is indefinite
    - m_k^Q is possibly unbounded below if B_k is only positive semi-definite

★ Second order approximation only works well in a small region / neighborhood

Definition 1.1 (trust-region subproblem)

The trust-region subproblem that we consider at the kth iterate is

$$\underset{s \in \mathbb{R}^n}{\text{minimize}}\ m_k(s) = f_k + g_k^T s + \tfrac{1}{2} s^T B_k s \quad \text{subject to} \quad \|s\| \le \delta_k$$

where B_k is a symmetric matrix and $\delta_k > 0$ is the trust-region radius.

Comments

- $B_k = H_k$ is always allowed! Not true for linesearch methods.
- if second derivatives are not available, then the symmetric rank-1 (SR1) quasi-Newton update is a very attractive option
  - may produce indefinite approximations B_k to the exact second derivative matrix H_k.
  - under certain assumptions it follows that

$$\lim_{k \to \infty} \|B_k - H_k\| = 0$$

so that fast convergence (generally) is recovered.
  - see quasi-Newton material discussed in the linesearch lecture slides for more details.

Trust-region methods overcome these difficulties by using a trust-region constraint

$$\|s\| \le \delta_k$$

for some "suitable" trust-region radius $\delta_k > 0$ and approximately solve

$$s_k = \underset{s \in \mathbb{R}^n}{\text{argmin}}\ m_k(s) \quad \text{subject to} \quad \|s\| \le \delta_k$$

- m_k may be any "reasonable" model of $f(x_k + s)$, e.g., m_k^L and m_k^Q
- global convergence results do not depend on the norm $\|\cdot\|$ that is used
- practical performance may depend on the norm $\|\cdot\|$ used!
- for simplicity, we focus on the second-order (Newton-like) quadratic model

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

and the ℓ_2 trust-region norm $\|\cdot\| = \|\cdot\|_2$
- other common trust-region norms simply add extra constants in the analysis
  - $\|x\|_2 \le \|x\|_1 \le \sqrt{n}\|x\|_2$
  - $\|x\|_\infty \le \|x\|_2 \le \sqrt{n}\|x\|_\infty$
  - $\|x\|_\infty \le \|x\|_1 \le n\|x\|_\infty$

Eigenvalue represents the property of the function

Algorithm 1 Almost a complete trust-region algorithm

1: Input an initial guess x_0.
2: Choose $0 < \gamma_d < 1 < \gamma_i$ and $0 < \eta_s \le \eta_{vs} < 1$.
3: Set $k \leftarrow 0$. Set $\delta_0 = 1$.
4: **loop**
5: Build the second-order model $m_k(s)$ of $f(x_k + s)$.
6: Approximately solve trust-region subproblem for s_k satisfying $m_k(s_k)$ "$<$" $m_k(0)$.
7: Set

$$\rho_k \leftarrow \frac{f_k - f(x_k + s_k)}{m_k(0) - m_k(s_k)} \equiv \frac{f_k - f(x_k + s_k)}{f_k - m_k(s_k)}.$$

Actual decrease / model decrease

8: **if** $\rho_k \ge \eta_{vs}$, **then**
9: Set $x_{k+1} \leftarrow x_k + s_k$ and $\delta_{k+1} \leftarrow \gamma_i \delta_k$. ▷ very successful
increase region (good trust region)
10: **else if** $\rho_k \ge \eta_s$, **then**
11: Set $x_{k+1} \leftarrow x_k + s_k$ and $\delta_{k+1} \leftarrow \delta_k$. ▷ successful
12: **else**
Not update
13: Set $x_{k+1} \leftarrow x_k$ and $\delta_{k+1} \leftarrow \gamma_d \delta_k$. ▷ unsuccessful
decrease region (bad trust region)
14: **end if**
15: Set $k \leftarrow k + 1$.
16: **end loop**

We desire a very weak requirement for approximately solving the trust-region subproblem

$$\underset{s\in\mathbb{R}^n}{\text{minimize}}\ \ m_k(s)\quad \text{subject to}\quad \|s\|\le\delta_k$$

- aim to achieve as much reduction in m_k as would an iteration of steepest descent.
- leads to the Cauchy point

$$s_k^C=-\alpha_k^C g_k$$

where

$$\begin{aligned}\alpha_k^C&=\underset{\alpha\ge 0}{\text{argmin}}\ \ m_k(-\alpha g_k)\quad \text{subject to}\quad \alpha\|g_k\|\le\delta_k\\&=\underset{0\le\alpha\le\delta_k/\|g_k\|}{\text{argmin}}\ \ m_k(-\alpha g_k)\end{aligned}$$

- explicit computation of α_k^C is very easy!.....and coming soon.
- we then allow for any step s_k that is "just as good" as the Cauchy step, i.e., satisfies

$$m_k(s_k)\le m_k(s_k^C)\quad\text{and}\quad\|s_k\|\le\delta_k$$

- in practice, we hope to do far better than just the Cauchy step.

Cauchy point is like the last choice at least

Definition 1.2 (decrease in the model)

Let $m_k(s)=m_k^Q(s)$ be the second-order model. For any step s, we define the decrease in the model m_k achieved by the step s as

$$\Delta m_k(s)\stackrel{\text{def}}{=}m_k(0)-m_k(s)$$

Lemma 1.3 (Achievable model decrease)

If $m_k(s)=m_k^Q(s)$ is the second-order model and s_k^C the Cauchy point, then

$$\Delta m_k(s_k^C)\ge\tfrac{1}{2}\|g_k\|\min\left[\frac{\|g_k\|}{\|B_k\|},\delta_k\right]\ge 0$$

Proof: Observe that

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

The result is immediate if $g_k=0$.

Therefore, we assume for the remainder that $g_k\ne 0$ and consider 2 cases related to the curvature along the steepest descent direction.

Case 1: $g_k^T B_k g_k\le 0$ Concave

In this case $m_k(-\alpha g_k)$ is unbounded below as α increases, which implies that the Cauchy point occurs on the trust-region boundary.

Equation (1), $g_k^T B_k g_k\le 0$, and $\alpha\ge 0$ imply that

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

Since the Cauchy point lies on the boundary of the trust region, we know that

$$\alpha_k^C=\frac{\delta_k}{\|g_k\|}$$

which may be combined with (2) to give

$$m_k(s_k^C)=m_k(-\alpha_k^C g_k)\le f_k-\alpha_k^C\|g_k\|^2=f_k-\delta_k\|g_k\|$$

which may then be combined with the definition of Δm_k to conclude that

$$\begin{aligned}\Delta m_k(s_k^C)&=m_k(0)-m_k(s_k^C)\\&=f_k-m_k(s_k^C)\\&\ge\delta_k\|g_k\|\ge\tfrac{1}{2}\delta_k\|g_k\|.\end{aligned}$$

Case 2: $g_k^T B_k g_k>0$

The minimizer of $m_k(-\alpha g_k)$ (disregarding the trust-region constraint) satisfies

$$\alpha_k^*\stackrel{\text{def}}{=}\underset{\alpha\ge 0}{\text{argmin}}\ \ m_k(-\alpha g_k)\equiv f_k-\alpha\|g_k\|^2+\tfrac{1}{2}\alpha^2 g_k^T B_k g_k \tag{3}$$

so that

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

We now consider to subcases.

Subcase 2a: $\alpha_k^*\ge\delta_k/\|g_k\|$

It must follow that α_k^C lies on the boundary and, therefore, that

$$\alpha_k^*=\frac{\|g_k\|^2}{g_k^T B_k g_k}\ge\frac{\delta_k}{\|g_k\|}=\alpha_k^C \tag{4}$$

which (after rearrangement) implies that

$$\alpha_k^C g_k^T B_k g_k\le\|g_k\|^2. \tag{5}$$

It now follows from (3), (5), and (4) that

$$\begin{aligned}\Delta m_k(s_k^C)&=m_k(0)-m_k(s_k^C)=f_k-m_k(-\alpha_k^C g_k)\\&=\alpha_k^C\|g_k\|^2-\tfrac{1}{2}[\alpha_k^C]^2 g_k^T B_k g_k\ge\tfrac{1}{2}\alpha_k^C\|g_k\|^2\\&=\tfrac{1}{2}\|g_k\|^2\frac{\delta_k}{\|g_k\|}=\tfrac{1}{2}\|g_k\|\delta_k.\end{aligned}$$

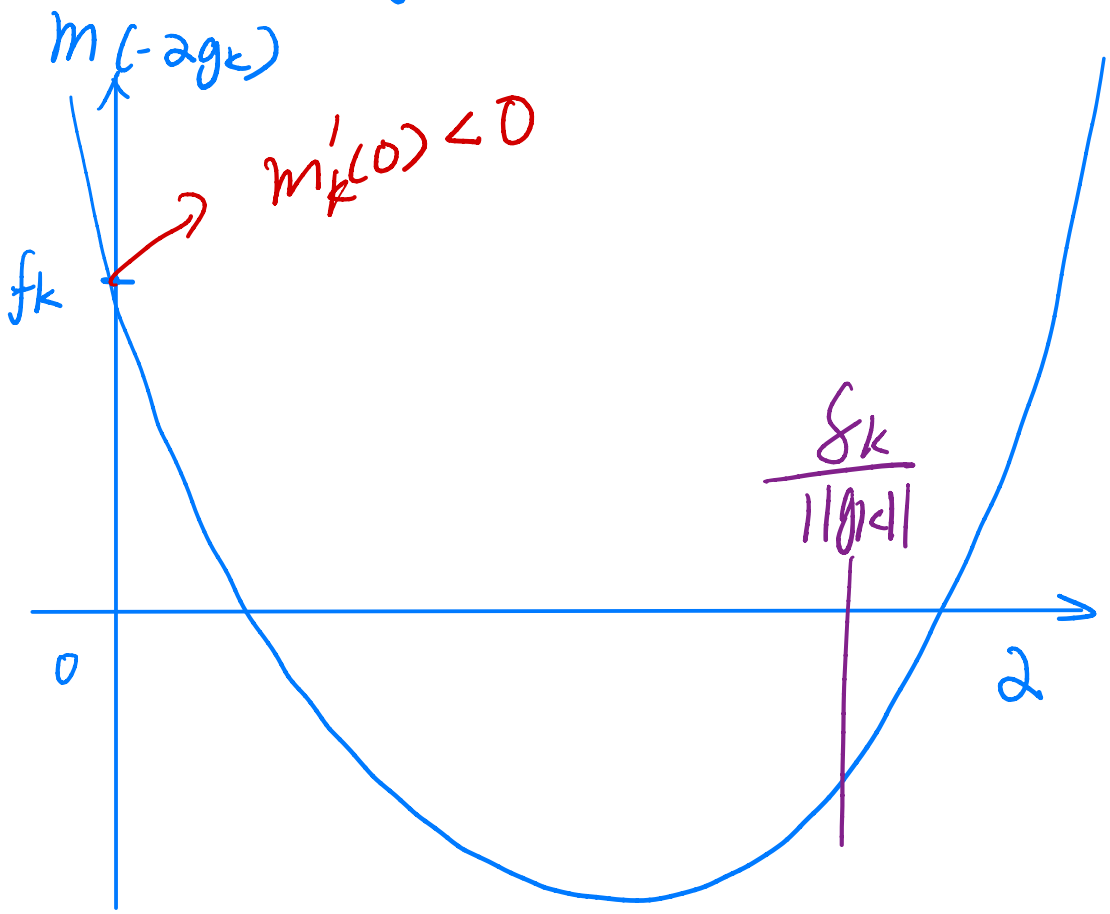
$$m(-\alpha g_k) = f_k - \alpha \|g_k\|^2 + \frac{1}{2}\alpha^2 g_k^T B g_k$$

quadratic function

Trust-region constrain

$$\|\alpha g_k\| \le \delta_k$$

$$\alpha \le \frac{\delta_k}{\|g_k\|}$$



$$\min_{0 \le \alpha \le \frac{\delta_k}{\|g_k\|}} m_k(-\alpha g_k)$$

Three Case:

① Convex $0 \le \alpha_k^c \le \frac{\delta_k}{\|g_k\|}$

② Convex $\alpha_k^c > \frac{\delta_k}{\|g_k\|} \ge 0$

③ Concave $\alpha_k^c < 0$

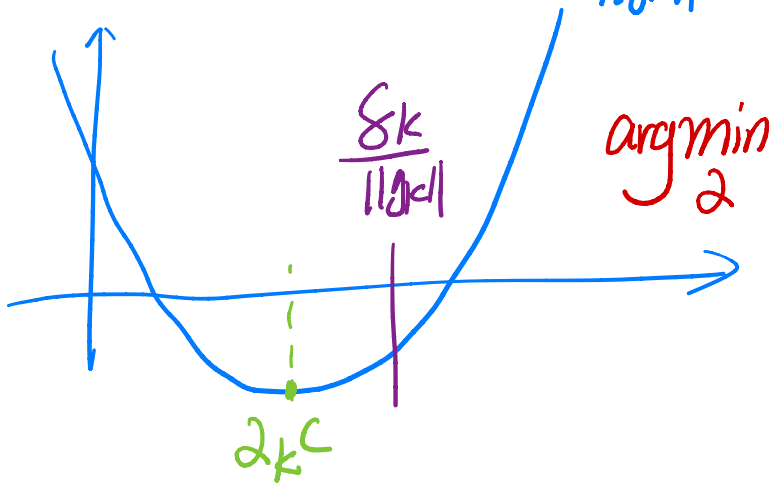
Three Case:

① Convex $0 \le \alpha_k^c \le \frac{\delta_k}{\|g_k\|}$

② Convex $\alpha_k^c > \frac{\delta_k}{\|g_k\|} \ge 0$

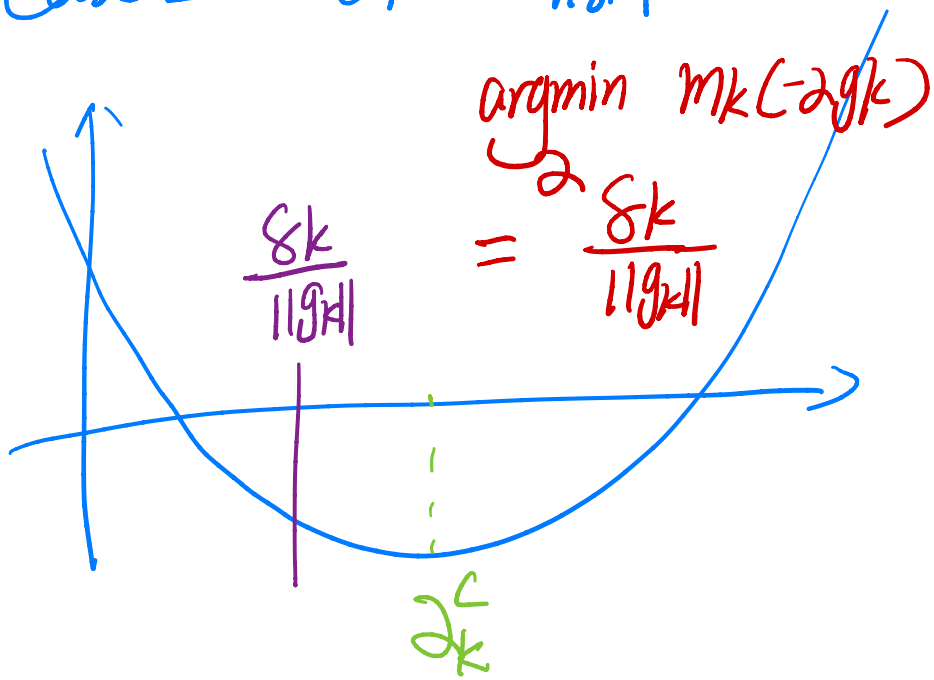
③ Concave $\alpha_k^c < 0$

Case 1: $0 \le \alpha_k^c \le \frac{\delta_k}{\|g_k\|}$ Convex



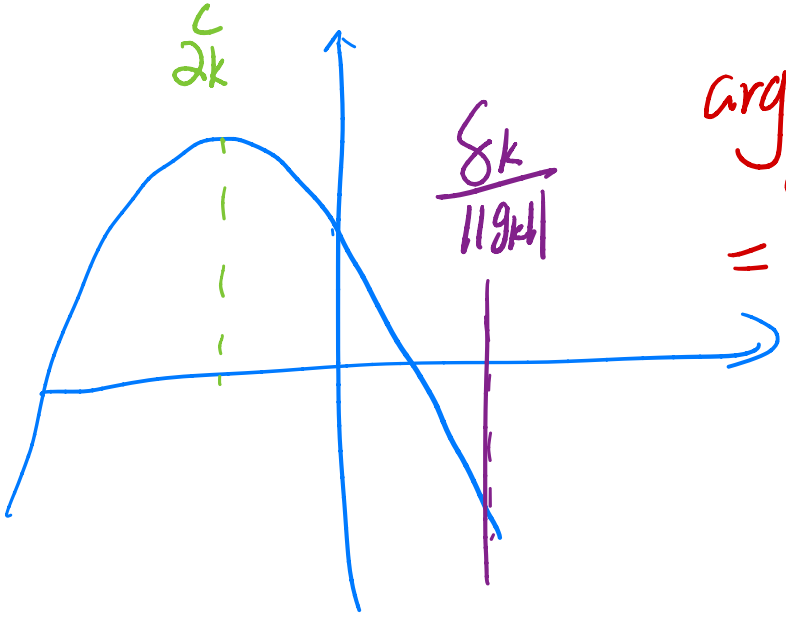
$\underset{\alpha}{\operatorname{argmin}}\ m_k(-\alpha g_k) = \alpha_k^c$

Case 2: $\alpha_k^c > \frac{\delta_k}{\|g_k\|} \ge 0$ Convex



$\underset{\alpha}{\operatorname{argmin}}\ m_k(-\alpha g_k) = \frac{\delta_k}{\|g_k\|}$

Case 3: Concave $\alpha_k^c < 0$



$\underset{\alpha}{\operatorname{argmin}}\ m_k(-\alpha g_k) = \frac{\delta_k}{\|g_k\|}$

Subcase 2b: $\alpha_k^* < \delta_k / \|g_k\|$

It follows that

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

so that

$$\begin{aligned}
\Delta m_k(s_k^C) &= m_k(0) - m_k(s_k^C) \\
&= f_k - m_k(-\alpha_k^C g_k) \\
&= \alpha_k^* \|g_k\|^2 - \tfrac{1}{2}(\alpha_k^*)^2 g_k^T B_k g_k \\
&= \frac{\|g_k\|^4}{g_k^T B_k g_k} - \tfrac{1}{2}\frac{\|g_k\|^4}{g_k^T B_k g_k} \\
&= \tfrac{1}{2}\frac{\|g_k\|^4}{g_k^T B_k g_k} \\
&\geq \tfrac{1}{2}\frac{\|g_k\|^2}{\|B_k\|}
\end{aligned}$$

where

$$|g_k^T B_k g_k| \leq \|g_k\|^2 \|B_k\|$$

follows from the Cauchy-Schwartz and matrix norm inequalities. ∎

Algorithm 2 A general trust-region algorithm

1: Input an initial guess x_0.
2: Choose $0 < \gamma_d < 1 < \gamma_i$ and $0 < \eta_s \leq \eta_{vs} < 1$.
3: Set $k \leftarrow 0$.
4: **loop**
5: Build the second-order model $m_k(s)$ of $f(x_k + s)$.
6: Find any trial step s_k that satisfies $\|s_k\| \leq \delta_k$ and $m_k(s_k) \leq m_k(s_k^C)$.
7: Set

$$\rho_k \leftarrow \frac{f_k - f(x_k + s_k)}{m_k(0) - m_k(s_k)} \equiv \frac{f_k - f(x_k + s_k)}{\Delta m_k(s_k)}.$$

8: **if** $\rho_k \geq \eta_{vs}$, **then**
9: Set $x_{k+1} \leftarrow x_k + s_k$ and $\delta_{k+1} \leftarrow \gamma_i \delta_k$. ▷ very successful → Good approximation
10: **else if** $\rho_k \geq \eta_s$, **then**
11: Set $x_{k+1} \leftarrow x_k + s_k$ and $\delta_{k+1} \leftarrow \delta_k$. ▷ successful
12: **else**
13: Set $x_{k+1} \leftarrow x_k$ and $\delta_{k+1} \leftarrow \gamma_d \delta_k$. ▷ unsuccessful → Bad approximation
14: **end if**
15: Set $k \leftarrow k + 1$.
16: **end loop**

- In practice, one should use a termination test such as used:

$$\|g_k\| \leq 10^{-6} \max(1, \|g(x_0)\|)$$

- Typical values: $\eta_s = 0.1$, $\eta_{vs} = 0.9$, $\gamma_d = 1/2$, $\gamma_i = 2$

Corollary 1.4

If $m_k(s) = m_k^Q(s)$ is the second-order model, and s_k is an improvement on the Cauchy point within the trust-region, i.e.,

$$m_k(s_k) \leq m_k(s_k^C) \quad \text{and} \quad \|s_k\| \leq \delta_k,$$

then

$$\Delta m_k(s_k) \stackrel{\text{def}}{=} m_k(0) - m_k(s_k) \geq \tfrac{1}{2}\|g_k\| \min\left[\frac{\|g_k\|}{\|B_k\|}, \delta_k\right] \geq 0$$

Note:

- $\Delta m_k(s_k) \geq 0$
- if $\Delta m_k(s_k) = 0$, then $g_k = 0$
- if we are not first-order optimal, i.e., $g_k \neq 0$, then

$$\Delta m_k(s_k) > 0$$

Lemma 1.5 (Difference between the model and the objective)

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

Function Hessian — Model Hessian

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

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

Proof:

The Mean Value Theorem implies that — Taylor approximation ?

$$f(x_k + s_k) = f_k + g_k^T s_k + \tfrac{1}{2} s_k^T H(\xi_k) s_k$$

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

$$\begin{aligned}
|f(x_k + s_k) - m_k(s_k)| &= \tfrac{1}{2}|s_k^T H(\xi_k) s_k - s_k^T B_k s_k| \\
&\leq \tfrac{1}{2}|s_k^T H(\xi_k) s_k| + \tfrac{1}{2}|s_k^T B_k s_k| \\
&\leq \tfrac{1}{2}(\kappa_h + \kappa_b)\|s_k\|^2 \\
&\leq \kappa_d \delta_k^2
\end{aligned}$$

where we have used the triangle-inequality, the Cauchy-Schwartz inequality, and the fact that $\|s_k\| \leq \delta_k$. ∎

Lemma 1.6 (Progress at non-optimal points)

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

$$\delta_k \leq \|g_k\| \min\left(\frac{1}{\kappa_h + \kappa_b}, \frac{(1-\eta_{vs})}{2\kappa_d}\right). \tag{6}$$

Then iteration k is very successful and

$$\delta_{k+1} \geq \delta_k$$

If $\|g_k\|$ is big, unsuccessful iteration number is bounded

Proof:
By definition of κ_h and κ_b we know that

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

and then from (6) it follows that

$$\delta_k \leq \frac{\|g_k\|}{\kappa_h + \kappa_b} \leq \frac{\|g_k\|}{\|B_k\|}.$$

Combining this with Corollary 1.4 and the fact that $g_k \neq 0$ by assumption, yields

$$\Delta m_k(s_k) \geq \tfrac{1}{2}\|g_k\| \min\left[\frac{\|g_k\|}{\|B_k\|}, \delta_k\right] = \tfrac{1}{2}\|g_k\|\delta_k > 0.$$

$$\delta_k \leq \frac{\|g_k\|}{\|B_k\|}$$

Lemma 1.7 (Radius will not shrink to zero at non-optimal points)

Suppose that $f \in C^2$, that the true and model Hessians satisfy the bounds $\|H_k\| \leq \kappa_h$ and $\|B_k\| \leq \kappa_b$ for all k and some $\kappa_h \geq 1$ and $\kappa_b \geq 0$, and that $\kappa_d = \frac{1}{2}(\kappa_h + \kappa_b)$. Suppose furthermore that there exists a constant ϵ and $k_0 \in \mathbb{N}$ such that $\|g_k\| \geq \epsilon > 0$ for all $k \geq k_0$. Then

$$\delta_k \geq \delta_{min} \stackrel{\text{def}}{=} \epsilon\gamma_d \min\left(\frac{1}{\kappa_h + \kappa_b}, \frac{(1-\eta_{vs})}{2\kappa_d}\right) > 0$$

for all $k \geq k_1$ for some $k_1 \in \mathbb{N}$.

Proof:
We first observe that if there is some $k' \geq k_0$ such that $\delta_{k'} \geq \epsilon \min\left(\frac{1}{\kappa_h+\kappa_b}, \frac{(1-\eta_{vs})}{2\kappa_d}\right)$, then for all $k \geq k'$ we must have $\delta_k \geq \delta_{\min} \stackrel{\text{def}}{=} \epsilon\gamma_d \min\left(\frac{1}{\kappa_h+\kappa_b}, \frac{(1-\eta_{vs})}{2\kappa_d}\right)$. Indeed, suppose otherwise that $k \geq k'$ is the first iteration such that

$$\delta_k \geq \delta_{\min} > \delta_{k+1} = \gamma_d \delta_k. \tag{7}$$

Dividing the previous line by γ_d and using the fact that $\epsilon \leq \|g_k\|$ gives

$$\delta_k = \frac{\delta_{k+1}}{\gamma_d} \leq \frac{\delta_{\min}}{\gamma_d} = \epsilon \min\left(\frac{1}{\kappa_h + \kappa_b}, \frac{(1-\eta_{vs})}{2\kappa_d}\right) \leq \|g_k\| \min\left(\frac{1}{\kappa_h + \kappa_b}, \frac{(1-\eta_{vs})}{2\kappa_d}\right)$$

It then follows from Lemma 1.6 that $\delta_{k+1} \geq \delta_k$, which contradicts (7).

From the previous slide we had

$$\Delta m_k(s_k) \geq \tfrac{1}{2}\|g_k\|\delta_k > 0$$

which may now be combined with the definition of ρ_k in line 7 of Algorithm 2, Lemma 1.5, and (6) to deduce that

$$\begin{aligned}
|\rho_k - 1| &= \left|\frac{f(x_k + s_k) - m_k(s_k)}{m_k(0) - m_k(s_k)}\right| \\
&= \frac{|f(x_k + s_k) - m_k(s_k)|}{\Delta m_k(s_k)} \\
&\leq 2\frac{\kappa_d \delta_k^2}{\|g_k\|\delta_k} \\
&= 2\frac{\kappa_d \delta_k}{\|g_k\|} \\
&\leq 1 - \eta_{vs}
\end{aligned}$$

$m_k(0) = f(x_k)$

$\leq \kappa_d \delta_k^2$

$\geq \frac{1}{2}\|g_k\|\delta_k$

$\delta_k \leq \|g_k\| \cdot \frac{(1-\eta_{vs})}{2\kappa_d}$

which implies that $\rho_k \geq \eta_{vs}$ and that the iteration is very successful. It now follows from line 9 of Algorithm 2 that

$$\delta_{k+1} \geq \delta_k$$

$|\rho_k - 1| \leq 1 - \eta_{vs}$

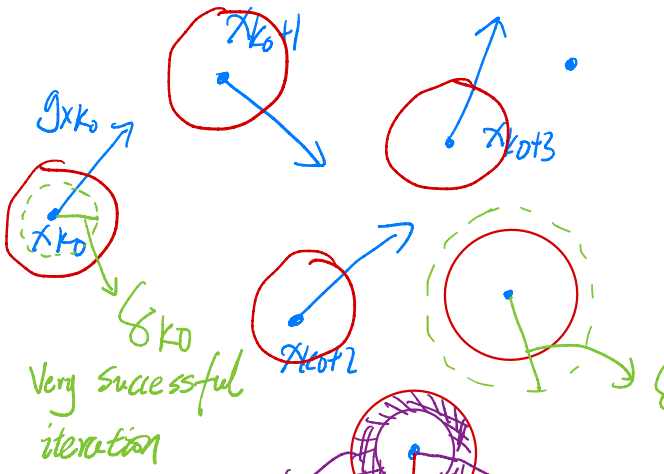
which is the desired result. ∎

We now have to show that there exists some $k' \geq k_0$ such that $\delta_{k'} \geq \epsilon \min\left(\frac{1}{\kappa_h+\kappa_b}, \frac{(1-\eta_{vs})}{2\kappa_d}\right)$. But this is true because whenever we have an iteration such that $\delta_{k'} < \epsilon \min\left(\frac{1}{\kappa_h+\kappa_b}, \frac{(1-\eta_{vs})}{2\kappa_d}\right)$, we have a very successful iteration by Lemma 1.6, and therefore, we strictly increase the radius by the factor $\gamma_i > 1$, i.e., $\delta_{k+1} = \gamma_i \delta_k$. ∎

lemma 1.6 → $\delta_k \le \|g_k\| \min\left(\underbrace{\frac{1}{\kappa_b + \kappa_h}}_{c}, \frac{1-\eta_{vs}}{2\kappa_d}\right) \Rightarrow \delta_k \le c\|g_k\|$

lemma 1.7

iterates is very successful



$g_{x_k} > \varepsilon$

$c\|g_{x_k}\| > c\varepsilon$

all red circle is $c\varepsilon$

↳ not guarantee for very successful iteration

Even with unsuccessful iteration, δ_k decreus by γ_d it cannot go below that region

$\forall k \ge k'$ → first time $\delta_{k'} > c\varepsilon$

$\delta_k \ge \gamma_d c\varepsilon$ → The worst case (smallest δ_k) for it is $\delta_k = c\varepsilon$ → not successful (might)

Then by algorithm, $\delta_{k+1} = \gamma_d c\varepsilon$

it cannot go beyond $\gamma_d c\varepsilon$.

$0 < \gamma_d < 1$

Once $\delta_k < c\varepsilon$ → $\delta_{k+1} = \gamma_i c\varepsilon$

$1 < \gamma_i$

Lemma 1.8 (Possible finite termination)

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 very successful and successful iterations. Then $x_k = x_*$ for all sufficiently large k and $g(x_*) = 0$.

Proof:
From the assumptions of this lemma, it follows that there exists some x_* such that

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

for all $j \geq 1$, where k_0 is the index of the last successful iterate.

Since all iterations are unsuccessful for k sufficiently large, we also know that

$$\lim_{k\to\infty} \delta_k = 0 \tag{8}$$

If $g(x_{k_0+1}) \neq 0$, let $\epsilon = \|g(x_{k_0+1})\| > 0$. By Lemma 1.7

$$\delta_k \geq \delta_{\min} \stackrel{\text{def}}{=} \epsilon\gamma_d \min\left(\frac{1}{\kappa_h + \kappa_b}, \frac{(1-\eta_{vs})}{2\kappa_d}\right),$$

contradicting (8). Thus, we must conclude that

$$g(x_*) = g(x_{k_0+1}) = 0$$

which is the desired result. ∎

To prove that outcome 3 must occur, we assume (to reach a contradiction) that there exists $\epsilon > 0$ and $k_0 \in \mathbb{N}$ such that

$$\|g_k\| \geq \epsilon > 0 \quad \text{for all } k \geq k_0. \tag{9}$$

It follows from the definition of $\mathcal{S}$, the definition of ρ_k in line 7 of Algorithm 2, Corollary 1.4, (9), and Lemma 1.7 that

$$f_k - f_{k+1} \geq \eta_s \Delta m_k(s_k) \geq \tfrac{1}{2}\eta_s \epsilon \min\left[\frac{\epsilon}{\kappa_b}, \delta_{\min}\right] \stackrel{\text{def}}{=} \delta_\epsilon > 0 \quad \text{for all } k \in \mathcal{S} \text{ such that } k \geq k_0$$

Picking $j \geq 1$ and then summing over all $k \in \mathcal{S}$ such that $k \leq j$ gives

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

Taking the limit as $j \to \infty$ gives

$$\lim_{j\to\infty} (f_0 - f_{j+1}) \geq \lim_{j\to\infty} \sum_{\substack{k=0 \\ k\in\mathcal{S}}}^{j} \delta_\epsilon = \sum_{\substack{k=0 \\ k\in\mathcal{S}}}^{\infty} \delta_\epsilon = \infty$$

which implies that f is unbounded below. This contradicts our assumption and, therefore, we must conclude that (9) is false, which means that there exists a subsequence of the gradients that converges to zero, i.e.,

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

∎

Theorem 1.9 (Global convergence of some subsequence)

Suppose that

- $f \in C^2$
- *true and model Hessians satisfy the bounds $\|H_k\| \leq \kappa_h$ and $\|B_k\| \leq \kappa_b$ for all k and some $\kappa_h \geq 1$ and $\kappa_b \geq 0$*

Then one of the following must occur:

1. *finite termination, i.e., there exists some finite k such that*
$$g_k = 0$$
2. *unbounded objective function, i.e.,*
$$\lim_{k\to\infty} f_k = -\infty$$
3. *convergence of a subsequence of the gradients, i.e.,*
$$\liminf_{k\to\infty} \|g_k\| = 0$$

Proof:
Let $\mathcal{S}$ be the index set of successful and very successful iterations, i.e.,

$$\mathcal{S} = \{k : x_{k+1} \leftarrow x_k + s_k\}.$$

Lemma 1.8 implies that outcome 1 is true when $|\mathcal{S}| < \infty$. Therefore, for the remainder, we assume that $|\mathcal{S}| = \infty$ and that f_k is bounded below (otherwise outcome 2 holds).

Theorem 1.10 (Global convergence)

Suppose that

- $f \in C^2$
- *true and model Hessians satisfy the bounds $\|H_k\| \leq \kappa_h$ and $\|B_k\| \leq \kappa_b$ for all k and some $\kappa_h \geq 1$ and $\kappa_b \geq 0$*

Then one of the following must occur:

1. *finite termination, i.e., there exists some finite k such that*
$$g_k = 0$$
2. *unbounded objective function, i.e.,*
$$\lim_{k\to\infty} f_k = -\infty$$
3. *convergence of the gradients, i.e.,*
$$\lim_{k\to\infty} g_k = 0$$

What is the difference between this theorem and the previous theorem?

- Case 3 is now a limit instead of a liminf.

Proof:
Suppose that outcome 1 and outcome 2 do not occur, i.e., that $g_k \neq 0$ for all $k \geq 0$ and f_k is bounded from below. We now wish to show that outcome 3 must occur.

For a proof by contradiction, assume that there is an $\epsilon > 0$ and a subsequence $\{t_i\} \subseteq \mathcal{S}$ such that

$$\|g_{t_i}\| \geq 2\epsilon > 0 \quad \text{for some } \epsilon \text{ and all } i. \tag{10}$$

On the other-hand, Theorem 1.9 implies that there exists a sequence $\{\ell_i\} \subseteq \mathcal{S}$ such that

$$\|g_k\| \geq \epsilon \quad \text{for} \quad t_i \leq k < \ell_i \quad \text{and} \quad \|g_{\ell_i}\| < \epsilon. \tag{11}$$

We now restrict our attention to indices in the set

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

Note from (11) that

$$\|g_k\| \geq \epsilon \quad \text{for all } k \in \mathcal{K}. \tag{12}$$

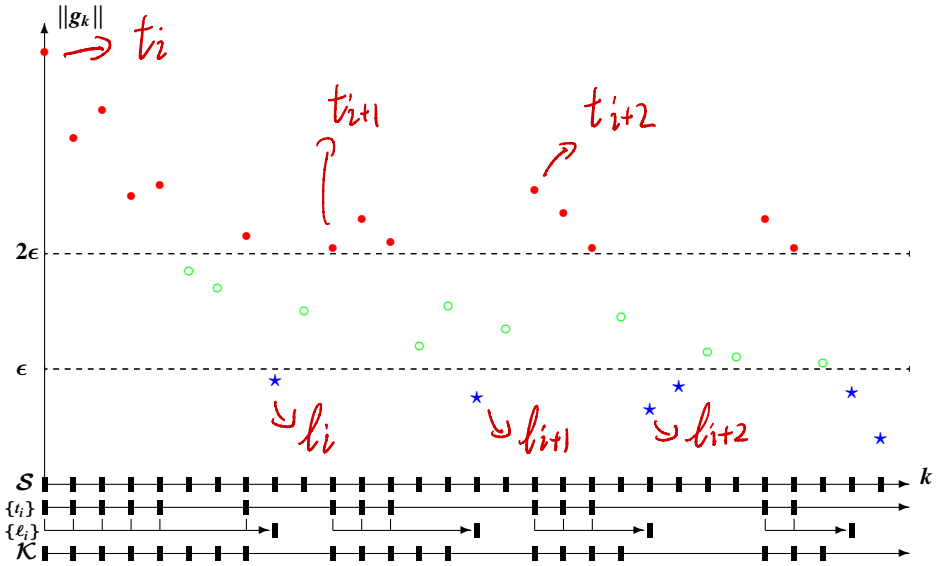


Figure: The subsequences in the proof of Theorem 1.10.

As in proof of Theorem 1.9, (12) implies

$$f_k - f_{k+1} \geq \eta_s \Delta m_k(s_k) \geq \tfrac{1}{2}\eta_s \epsilon \min\left[\frac{\epsilon}{\kappa_b}, \delta_k\right] \quad \text{for all } k \in \mathcal{K}. \tag{13}$$

Since the LHS of (13) converges to 0 as $k \to \infty$ (why?), we may conclude that

$$\delta_k \leq \frac{2}{\epsilon\eta_s}[f_k - f_{k+1}] \quad \text{for } k \in \mathcal{K} \text{ sufficiently large.}$$

We may now use the triangle-inequality, the definition of the trust-region radius δ_k, and the previous inequality to conclude that

$$\|x_{t_i} - x_{\ell_i}\| \leq \sum_{\substack{j=t_i \\ j\in\mathcal{K}}}^{\ell_i - 1} \|x_j - x_{j+1}\| \leq \sum_{\substack{j=t_i \\ j\in\mathcal{K}}}^{\ell_i - 1} \delta_j \leq \frac{2}{\epsilon\eta_s}[f_{t_i} - f_{\ell_i}] \quad \text{for } i \text{ sufficiently large.} \tag{14}$$

Since the RHS of (14) converges to 0, we know that

$$\lim_{i\to\infty} \|x_{t_i} - x_{\ell_i}\| = 0.$$

Combining this with the assumed continuity of g implies that

$$\lim_{i\to\infty} \|g_{t_i} - g_{\ell_i}\| = 0.$$

However, this is a contradiction since $\|g_{t_i} - g_{\ell_i}\| \geq \epsilon$ by definition of $\{t_i\}$ and $\{\ell_i\}$. ∎

Convergence rates for trust region methods

- **Global Convergence:** With some tweaks to the algorithm, one can prove $O\left(\left(\frac{1}{\epsilon}\right)^2\right)$ (and even $O\left(\left(\frac{1}{\epsilon}\right)^{3/2}\right)$) convergence to ϵ-stationary points. See the following short paper for a recent, clean analysis:

  "Concise complexity analyses for trust region methods", Optimization Letters, 2018, DOI: 10.1007/s11590-018-1286-2 by F. Curtis, Z. Lubberts and D. Robinson.

- **Local Convergence:** If B_k is taken to be the Hessian and the model is solved to give steps that are better than the Cauchy step and are asymptotically similar to the Newton steps, then one can prove superlinear (and with stronger assumptions, even quadratic) local convergence. See Theorem 4.9 in Nocedal-Wright.

Recall:

For the quadratic model $m_k(s) = f_k + g_k^T s + \frac{1}{2} s^T B_k s$, the trust-region subproblem is

$$\underset{s \in \mathbb{R}^n}{\text{minimize}} \;\; m_k(s) \quad \text{subject to} \quad \|s\| \leq \delta_k$$

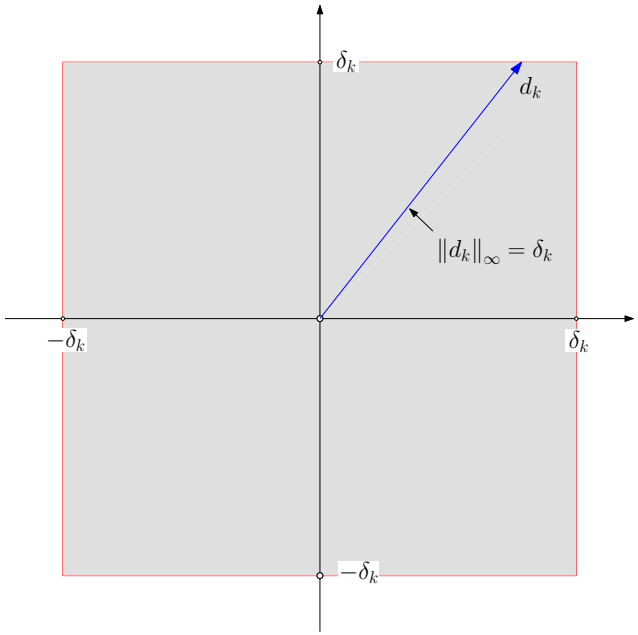
and minimizers of this model are also minimizers of

$$\underset{s \in \mathbb{R}^n}{\text{minimize}} \;\; g_k^T s + \frac{1}{2} s^T B_k s \quad \text{subject to} \quad \|s\| \leq \delta_k$$

Question: Which norm should we use?

Answer: Popular choices are the infinity norm and the two norm.

$$\{ d : \|d\|_\infty \leq \delta_k \}$$



The effect of different norms with small δ_k

The choice of norm determines the behavior of s_k as $\delta_k \to 0$.

If $\|s\| \ll 1$ then

$$f_k + g_k^T s + \frac{1}{2} s^T B_k s \approx f_k + g_k^T s$$

so that for $\delta_k \ll 1$ we have

$$\begin{array}{ll} \underset{s \in \mathbb{R}^n}{\text{minimize}} & m_k(s) = f_k + g_k^T s + \frac{1}{2} s^T B_k s \\ \text{subject to} & \|s\| \leq \delta_k \end{array} \quad \approx \quad \begin{array}{ll} \underset{s \in \mathbb{R}^n}{\text{minimize}} & f_k + g_k^T s \\ \text{subject to} & \|s\| \leq \delta_k \end{array}$$

$\Longrightarrow$ solution s_k approaches the steepest-descent direction of length δ_k as $\delta_k \to 0$.

$\Longrightarrow$ $s_k \to 0$ in the direction of the steepest-descent direction.

Examples:

- the two-norm
$$s_k \to -\frac{\delta_k}{\|g_k\|_2} g_k \quad \text{as} \quad \delta_k \to 0$$
- the infinity norm
$$s_k \to -\delta_k \widehat{e} \quad \text{with} \quad \widehat{e_j} = \text{sign}(g_j(x_k)) \quad \text{as} \quad \delta_k \to 0$$

corner

For the infinity norm, the subproblem is

$$\begin{array}{ll} \underset{s \in \mathbb{R}^n}{\text{minimize}} & f_k + g_k^T s + \frac{1}{2} s^T B_k s \\ \text{subject to} & -\delta_k e \leq s \leq \delta_k e \end{array}$$

where $e \in \mathbb{R}^n$ is a vector of all ones.

- a quadratic program (QP) and possibly nonconvex.
- local solutions may be computed using bound-constrained optimization algorithms.
- finding the global minimizer is NP-hard, i.e., it is a very hard problem!

Example 2.1 (minimizer inside the trust-region)

Consider

$$\min_s \quad m(s) = f + g^T s + \tfrac{1}{2} s^T B s \quad \text{subject to} \quad \|s\|_\infty \leq 4$$

with

$$f = 0, \quad g = \begin{pmatrix} 2 \\ 4 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$$

The unique global minimizer

$$s^* = \begin{pmatrix} -2 \\ -2 \end{pmatrix}$$

lies inside the trust region with $m(s^*) = -6$

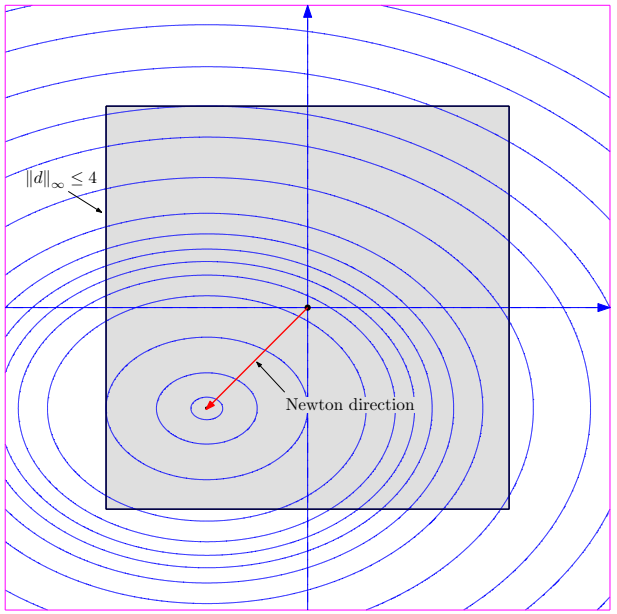


Figure: Plot associated with Example 2.1.

Example 2.2 (minimizer on the trust-region)

Consider

$$\min_s \quad m(s) = f + g^T s + \tfrac{1}{2} s^T B s \quad \text{subject to} \quad \|s\|_\infty \leq 4$$

with

$$f = 0, \quad g = \begin{pmatrix} 2 \\ 4 \end{pmatrix}, \quad \text{and} \quad B = \begin{pmatrix} 1 & 0 \\ 0 & -2 \end{pmatrix}$$

The model $m(s)$ is unbounded below and

$$s^N = \begin{pmatrix} -2 \\ 2 \end{pmatrix}$$

is the step to the saddle point of $m(s)$.

- $g^T s^N = 4 \implies s^N$ is not a descent direction for f (or m).
- The unique global minimizer lies on the boundary of the trust region.
- There are two local solutions.

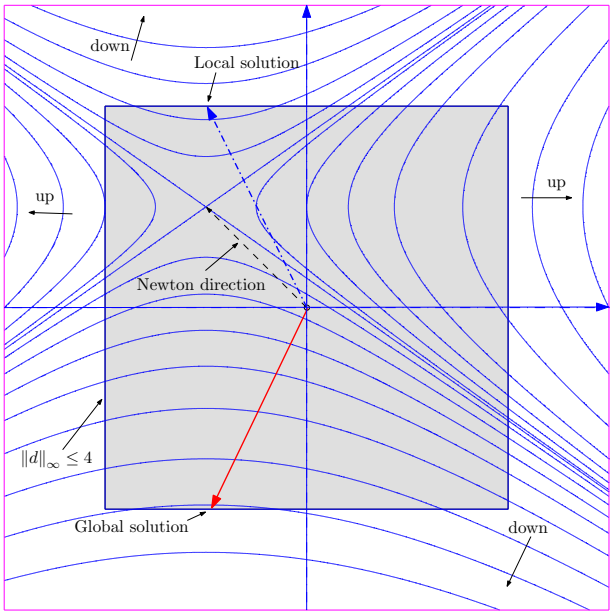


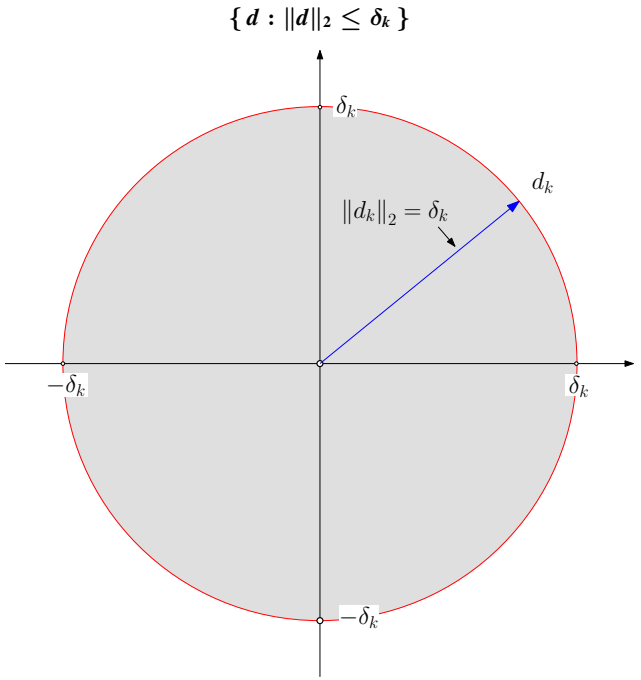
Figure: Plot associated with Example 2.2.

For the two-norm, the subproblem has the form

$$\begin{aligned} \underset{s\in\mathbb{R}^n}{\text{minimize}} \quad & m_k(s) = f_k + g_k^T s + \tfrac{1}{2} s^T B_k s \\ \text{subject to} \quad & \|s\|_2 \le \delta_k \end{aligned}$$

This is a nonlinearly constrained optimization problem.

- amazingly, there are efficient algorithms for computing the global minimizer
- we will focus on this two-norm trust-region subproblem



Example 2.3 (minimizer inside the trust-region)

Consider

$$\min_s \quad m(s) = f + g^T s + \tfrac{1}{2} s^T B s \quad \text{subject to} \quad \|s\|_2 \le 4$$

with

$$f = 0, \quad g = \begin{pmatrix} 2 \\ 4 \end{pmatrix}, \quad \text{and} \quad B = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$$

The unique global minimizer is the Newton step

$$s^* = \begin{pmatrix} -2 \\ -2 \end{pmatrix}$$

and since

$$\|s^*\|_2 = 2\sqrt{2}$$

lies inside the trust region with $m(s^*) = -6$.

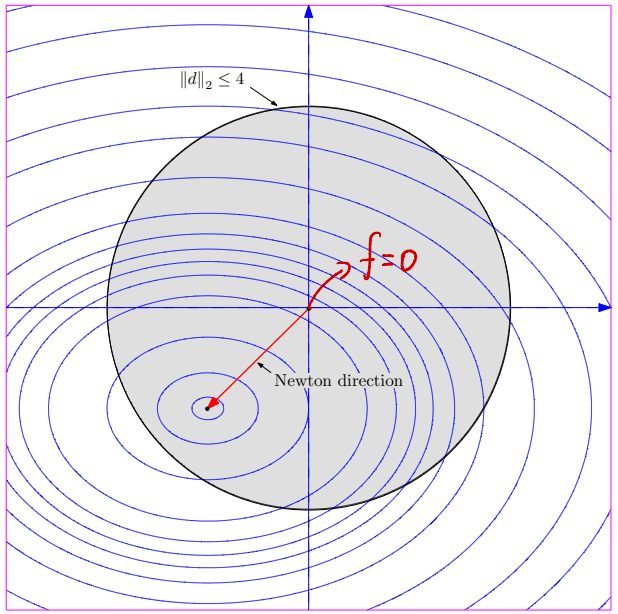


Figure: Plot associated with Example 2.3.

Example 2.4 (minimizer on the trust-region)

Consider

$$\min_s \; m(s) = f + g^T s + \tfrac{1}{2} s^T B s \quad \text{subject to} \quad \|s\|_2 \leq 4$$

with

$$f = 0, \quad g = \begin{pmatrix} 2 \\ 4 \end{pmatrix}, \quad \text{and} \quad B = \begin{pmatrix} 1 & 0 \\ 0 & -2 \end{pmatrix}$$

The unique global minimizer

$$s^* = \begin{pmatrix} -0.49902 \\ -3.96875 \end{pmatrix}$$

lies on the boundary of the trust region with

$$m(s^*) = -32.49962.$$

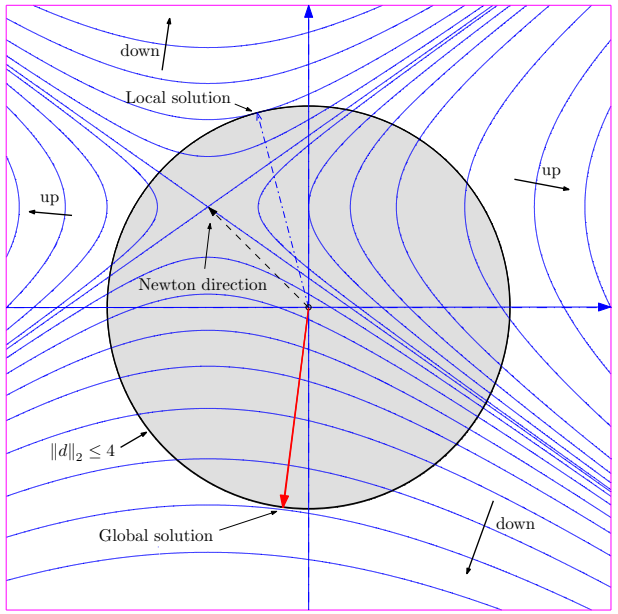


Figure: Plot associated with Example 2.4.

Observation:

The trust-region constraint

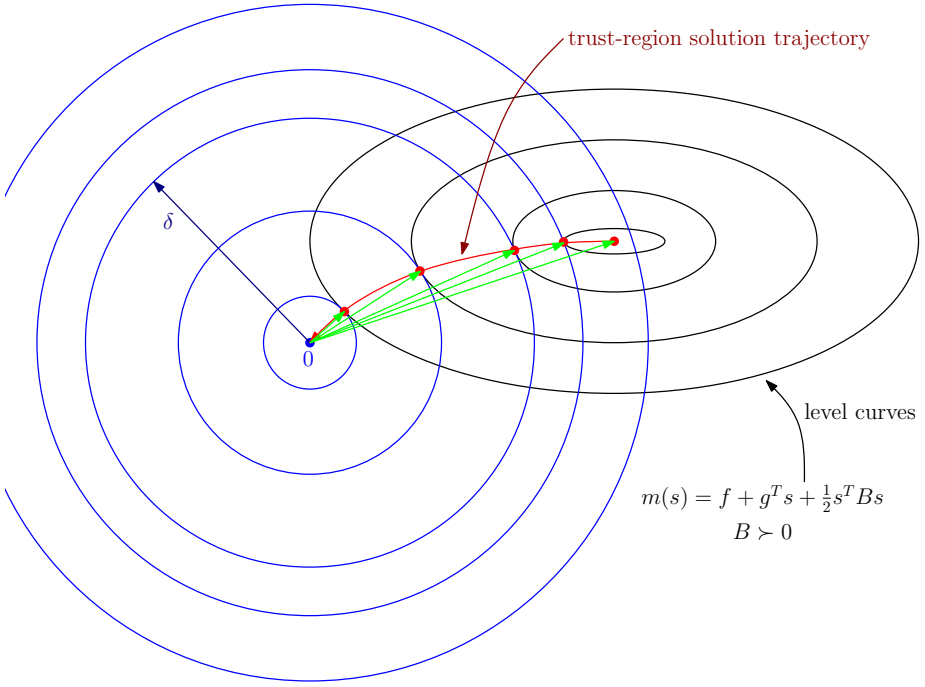
$$\|s\| \leq \delta$$

effects both the length and direction of the trial direction.

Consider the following trust-region subproblem

$$\begin{aligned} \underset{s \in \mathbb{R}^n}{\text{minimize}} \quad & m(s) = f + g^T s + \tfrac{1}{2} s^T B s \\ \text{subject to} \quad & \|s\|_2 \leq \delta \end{aligned}$$

- f, g, and B are fixed
- given a value of δ, let $s(\delta)$ denote the global minimizer (for simplicity assume it is unique)
- $s(\delta)$ for $\delta \geq 0$ traces out the trust-region solution trajectory



Goal: given a vector g and a symmetric (possibly indefinite) matrix B, find the global minimizer s_* to the problem

$$\underset{s\in\mathbb{R}^n}{\text{minimize}}\ \ m(s) \equiv f + s^T g + \tfrac{1}{2}s^T B s \quad \text{subject to} \quad \|s\|_2 \le \delta$$

i.e., find a vector s_* that satisfies

$$\|s_*\|_2 \le \delta \quad \text{and} \quad m(s_*) \le m(s) \ \text{ for all } \|s\|_2 \le \delta.$$

- "exact" solution implies Newton-like method
- Cauchy step s^C implies a steepest-descent-like method
- truncated linear CG (to be discussed later) lies somewhere in between

Observations:

- to prove global convergence of the trust-region method, we only require an approximate step s_A that satisfies

  $$m(s_A) \le m(s^C) \quad \text{and} \quad \|s_A\|_2 \le \delta$$

  since this implies that

  $$\Delta m(s_A) \ge \Delta m(s^C) \quad \text{(we proved convergence)}$$

- it is also sufficient to compute an approximate solution s_A that satisfies

  $$\Delta m(s_A) \ge \gamma \Delta m(s_*) \quad \text{and} \quad \|s_A\|_2 \le \delta \ \text{ for some } \gamma \in (0,1)$$

  since this implies that

  $$\Delta m(s_A) \ge \gamma \Delta m(s_*) = \gamma(m(0) - m(s_*)) \ge \gamma(m(0) - m(s^C)) = \gamma \Delta m(s^C)$$

From the previous slide, we know that when the trust-region constraint is "active" at the solution s_*, then there exists some $\lambda_* \ge 0$ such that

$$(B + \lambda_* I)s_* = -g \qquad (15)$$

local minimizer

Another observation:

The step s_* that satisfies (15) is also a solution to the unconstrained minimization problem

$$\underset{s\in\mathbb{R}^n}{\text{minimize}}\ \ f + g^T s + \tfrac{1}{2}s^T(B + \lambda_* I)s \longrightarrow \text{set the } \nabla\phi(s) = 0$$

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

for the unknown constant $\lambda_* \ge 0$ such that $(B + \lambda_* I) \succeq 0$.

- global minimizer s_* is unique if $(B + \lambda_* I) \succ 0$
- need to find λ_*

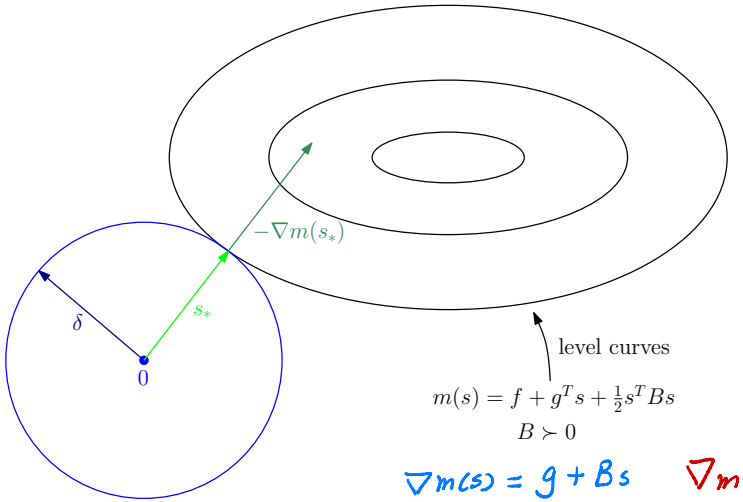
positive definite → convex function

Observation:

When the trust-region solution satisfies $\|s_*\|_2 = \delta$, then s_* is orthogonal to the level curve of $m(s)$ that runs through s_* ($-\nabla m(s_*)$ is orthogonal to sphere $\|s\|_2 = \delta$)

- in fact, s_* and $-\nabla m(s_*)$ point in the same direction
- there exists $\lambda_* \ge 0$ such that

$$\lambda_* s_* = -\nabla m(s_*) \quad \text{which is equivalent to} \quad (B + \lambda_* I)s_* = -g$$



$\nabla m(s) = g + Bs$ $\quad \nabla m(s_*) = -\lambda_* s_*$

$\nabla m(s_*) = g + Bs_*$ $\quad g + Bs_* = -\lambda_* s_*$

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

Theorem 2.5 (Characterization of the 2-norm trust-region solution)

A vector s_* is a global minimizer of

$$\underset{s\in\mathbb{R}^n}{\text{minimize}}\ \ m(s) \equiv f + s^T g + \tfrac{1}{2}s^T B s \quad \text{subject to} \quad \|s\|_2 \le \delta$$

if and only if $\|s_*\|_2 \le \delta$ and there exists a scalar λ_* such that

- $\lambda_* \ge 0$
- $(B + \lambda_* I)s_* = -g$
- $B + \lambda_* I$ *is positive semi-definite*
- $\lambda_*(\|s_*\|_2 - \delta) = 0$. → if $s_* \le \delta$, to achive this, only when $\lambda_* = 0$ (B is positive definite)

Moreover, if $B + \lambda_* I$ is positive definite, then s_* is unique.

- the "shifted" Hessian $B + \lambda_* I$ is positive semi-definite, which ensures that s_* is not an ascent direction. Is it guaranteed to be a descent direction?
- if $\|s_*\|_2 < \delta$, then $\lambda_* = 0$ and

  $$Bs^* = -g$$

  so that s^* is the Newton direction! (if $B = H$)
- if $\lambda_* > 0$, then $\|s_*\|_2 = \delta$ so that the global minimizer lies on the trust-region boundary.
- λ_* is actually a Lagrange multiplier for the problem

  $$\underset{s\in\mathbb{R}^n}{\text{minimize}}\ m(s) \quad \text{subject to} \quad \tfrac{1}{2}\|s\|_2^2 \le \tfrac{1}{2}\delta^2$$

  but this requires optimality conditions for constrained optimization (EN.553.762)

Example 2.6 (local versus global minimizer)

Consider

$$\min_d \; m(d) = f + g^T d + \tfrac{1}{2} d^T B d \quad \text{subject to} \quad \|d\|_2 \leq 4$$

where

$$f = 0, \quad g = \begin{pmatrix} 2 \\ 4 \end{pmatrix}, \quad \text{and} \quad B = \begin{pmatrix} 1 & 0 \\ 0 & -2 \end{pmatrix}.$$

In this case

$$(B + \lambda_* I) s_* = -g \quad \text{with} \quad \lambda_* = 3.00787$$

and that

$$B + \lambda_* I = \begin{pmatrix} 4.00787 & 0 \\ 0 & 1.00787 \end{pmatrix} \quad \text{is positive definite}$$

There is a local minimizer $\widehat{s}$ on the boundary of the trust region that satisfies

$$\widehat{s} = \begin{pmatrix} -1.0173 \\ 3.8684 \end{pmatrix} \quad \text{and} \quad m(\widehat{s}) = -1.0082.$$

In this case

$$(B + \widehat{\lambda} I) \widehat{s} = -g \quad \text{with} \quad \widehat{\lambda} = 0.9660$$

but

$$B + \widehat{\lambda} I = \begin{pmatrix} 1.9660 & 0 \\ 0 & -1.0340 \end{pmatrix} \quad \text{which is not positive definite}$$

and implies that $\widehat{s}$ and $\widehat{\lambda}$ do not satisfy the optimality conditions.

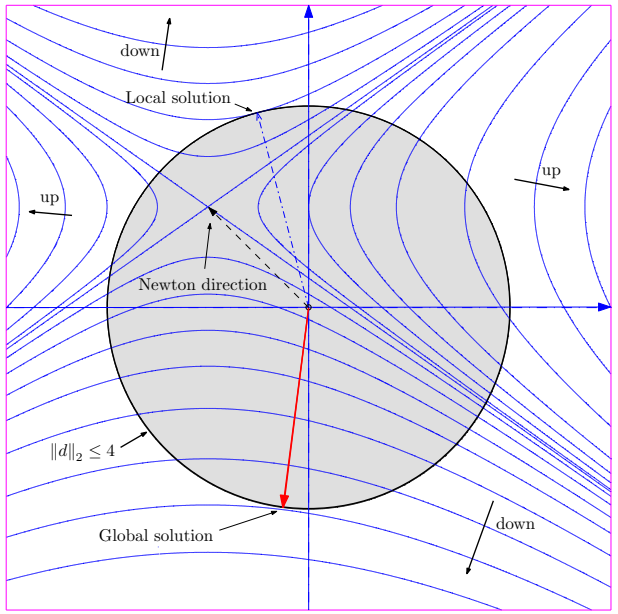


Figure: Plot associated with Example 2.6.

$$\underset{s \in \mathbb{R}^n}{\text{minimize}} \; m(s) = f + g^T s + \tfrac{1}{2} s^T B s \quad \text{subject to} \quad \|s\|_2 \leq \delta$$

symetric !!!

Three cases:

① check if B is psd
② if so, find s using $Bs = -g$
③ check if $s \leq B_\delta$ $\Rightarrow$ $s_* = s$ $\lambda_* = 0$

1. B positive-semidefinite and some vector s satisfies $Bs = -g$ and $\|s\|_2 \leq \delta$
   - $s_* = s$ and $\lambda^* = 0$
2. B positive-semidefinite and there does not exist a vector s that satisfies $Bs = -g$ and $\|s\|_2 \leq \delta$ s_* outside trust-region
   - $(B + \lambda_* I) s_* = -g$ and $\|s_*\|_2 = \delta$ → solve to get s_* and λ_*
   - nonlinear (quadratic) system in s and λ
3. B is not positive semi-definite (covers the indefinite case)
   - $(B + \lambda_* I) s_* = -g$ and $\|s_*\|_2 = \delta$
   - nonlinear (quadratic) system in s and λ

We will focus on cases 2 and 3.

Therefore, we only consider the case that $\|s_*\|_2 = \delta$.

Based on the characterization of a global minimizer as given by Theorem 2.5, we may describe our goal.

Goal: first version

Find a scalar $\lambda_* \geq 0$ and a vector s_* such that

$$B + \lambda_* I \succeq 0, \quad (B + \lambda_* I) s_* = -g, \quad \text{and} \quad \|s_*\|_2 = \delta$$

Consider the spectral decomposition $B = V \Lambda V^T$, where

$$V = \begin{pmatrix} v_1 & v_2 & \cdots & v_n \end{pmatrix} \quad \text{and} \quad \Lambda = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$$

with $B v_j = \lambda_j v_j$ for $j = 1, 2, \ldots, n$ and $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. It follows that

$$B + \lambda I = V \Lambda V^T + \lambda I = V(\Lambda + \lambda I) V^T$$

- $B + \lambda I \succeq 0$ if and only if $\lambda \geq -\lambda_n$
- $B + \lambda I \succ 0$ if $\lambda > -\lambda_n$

We find different s_* in term of λ_* then check if $\|s_*\|_2 = \delta$

Goal: second version

Find a scalar $\lambda_* \geq \max(0, -\lambda_n)$ and a vector s_* such that

$$(B + \lambda_* I) s_* = -g \quad \text{and} \quad \|s_*\|_2 = \delta$$

Definition 2.7

For $\lambda > -\lambda_n$ let $s(\lambda)$ denote the unique solution to the linear system

$$(B + \lambda I)s = -g$$

so that

$$(B + \lambda I)s(\lambda) = -g$$

If we define

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

then we are searching for a $\lambda_* \geq \max(0, -\lambda_n)$ such that

$$\psi(\lambda_*) = \delta$$

If $B + \lambda I$ is nonsingular, then $\det(B+\lambda I) \neq 0$

$$(B + \lambda I)^{-1} = V(\Lambda + \lambda I)^{-1}V^T$$

so that

$$s(\lambda) = -(B + \lambda I)^{-1}g = -\sum_{i=1}^{n} \frac{v_i^T g}{\lambda_i + \lambda} v_i$$

Taking norms and using the orthogonality of the columns of V, gives

$$\|s(\lambda)\|_2^2 = \|(B + \lambda I)^{-1}g\|_2^2 = \sum_{i=1}^{n} \frac{(v_i^T g)^2}{(\lambda_i + \lambda)^2}$$

which implies that $\psi(\lambda) = \|s(\lambda)\|_2$ has poles at $\lambda = -\lambda_i$ if $v_i^T g \neq 0$.

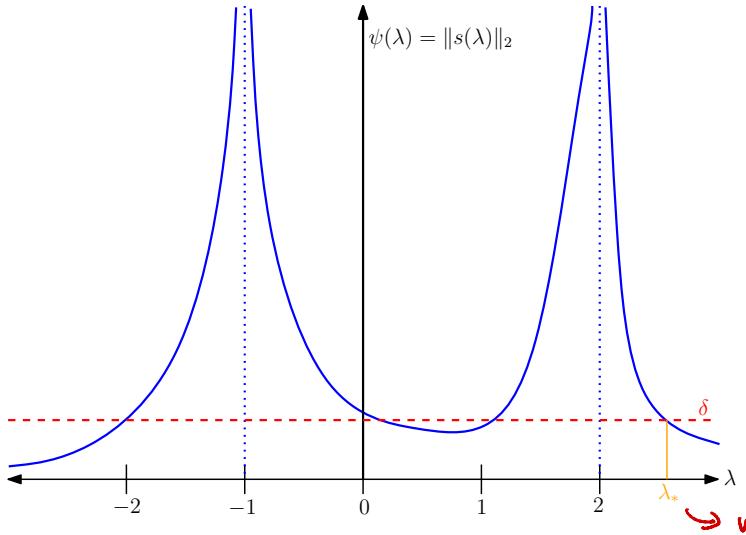


Figure: Plot representing $\|s(\lambda)\|_2$ associated with Example 2.8.

we want $\lambda_* > -\lambda_1$

$\Rightarrow \lambda_* > -(-2) = 2$

For this example it is clear that for any $\delta > 0$ there is a $\lambda_* > 2$ satisfying $\psi(\lambda_*) = \delta$ so that $s(\lambda_*)$ solves the trust-region subproblem.

Example 2.8 (poles of $\psi(\lambda) = \|s(\lambda)\|_2$)

Consider

$$\min_s \; f + g^T s + \tfrac{1}{2}s^T B s \quad \text{subject to} \quad \|s\|_2 \leq 4$$

with

$$f = 0, \qquad g = \begin{pmatrix} 2 \\ 4 \end{pmatrix}, \quad \text{and} \quad B = \begin{pmatrix} 1 & 0 \\ 0 & -2 \end{pmatrix}$$

It is easy to see that B has one positive and one negative eigenvalue, and that

$$s(\lambda) = \begin{pmatrix} -2/(\lambda + 1) \\ -4/(\lambda - 2) \end{pmatrix}$$

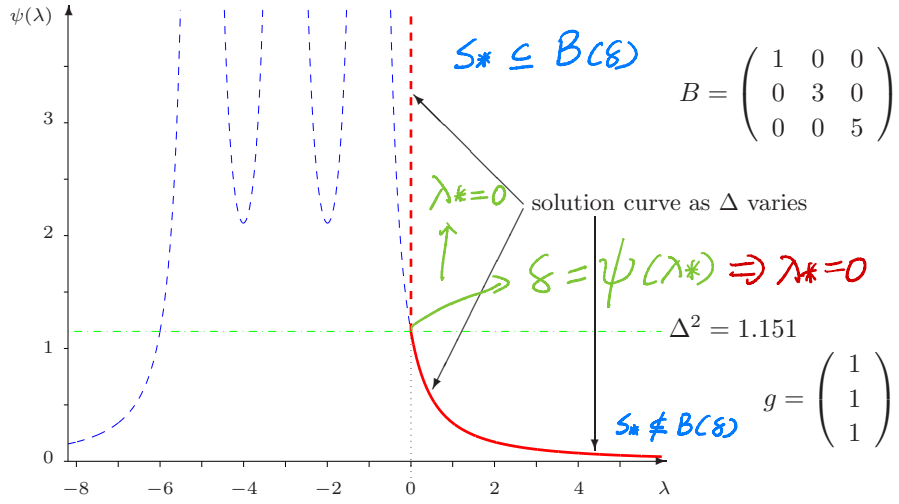
so that

$$\|s(\lambda)\|_2^2 = \frac{4}{(\lambda + 1)^2} + \frac{16}{(\lambda - 2)^2}.$$

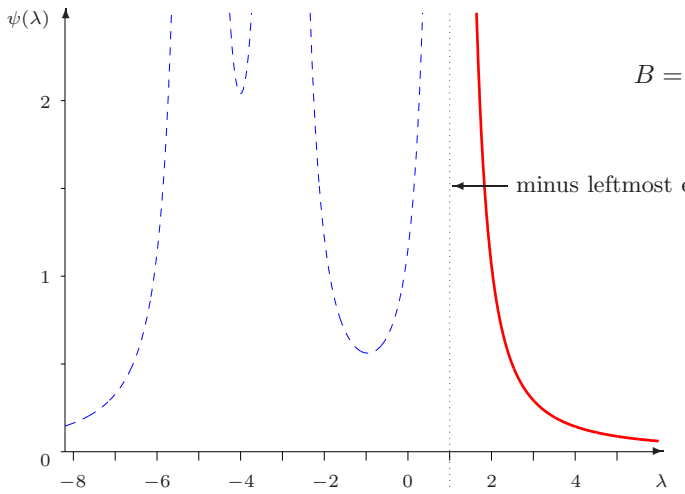
It follows that $\psi(\lambda) = \|s(\lambda)\|_2$ has two poles at

$$\lambda = -\lambda_1 = -1 \quad \text{and} \quad \lambda = -\lambda_2 = 2.$$

CONVEX EXAMPLE



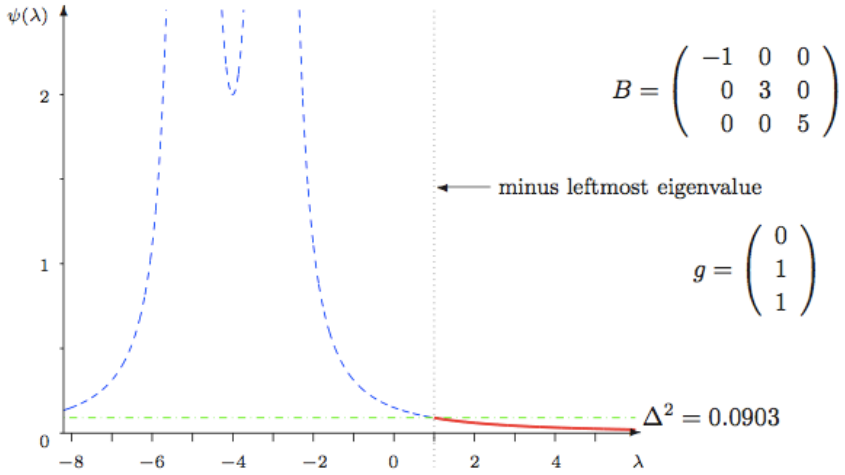
NONCONVEX EXAMPLE



$$B = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{pmatrix}$$

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

THE "HARD" CASE



$$B = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{pmatrix}$$

$$g = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$$

- $g^T v_3 = 0$ where $v_3 = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}^T$ so that there is no pole at $\lambda = -\lambda_3 = 1$.
- There are solutions for $\delta \leq 0.0903$.
- There are no obvious solutions for $\delta > 0.0903$ (of course there is one!)

Summary of the "hard case"

For indefinite B, the hard case occurs when g is orthogonal to the eigenvector v_n associated with the most negative eigenvalue λ_n

- OK if radius is "small enough"
- No "obvious" solution when radius is "big", but in fact a solution is of the form

$$s_{\lim} + \sigma v_n$$

where

$$s_{\lim} = \lim_{\lambda \downarrow -\lambda_n} s(\lambda)$$

and σ is chosen to satisfy

$$\|s_{\lim} + \sigma v_n\|_2 = \delta$$

Question: How do we actually solve $\|s(\lambda)\|_2 = \delta$?
Answer: We don't!....but instead solve the secular equation

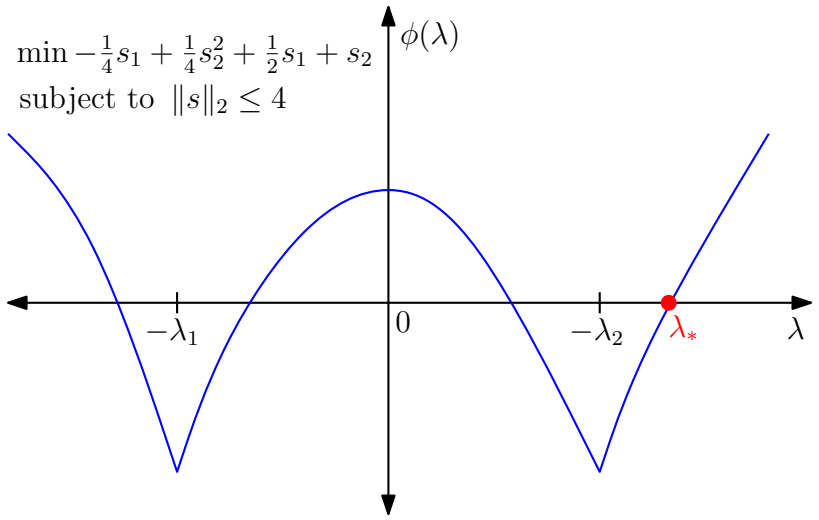
$$\phi(\lambda) \stackrel{\text{def}}{=} \frac{1}{\|s(\lambda)\|_2} - \frac{1}{\delta} = 0$$

using Newton's Method (safeguarded).

Properties of ϕ:

- no poles
- smallest at eigenvalues (except in hard case!)
- analytic function implies ideal for Newton
- usually nearly linear close to λ_*
- globally convergent (ultimately quadratic rate except in hard case)
- need to safeguard Newton's Method to deal with the hard case and when the solution is in the strict interior of the trust-region

The secular equation example



Algorithm 3 Newton's Method for solving the secular equation

1: Input symmetric matrix B, vector g, and trust-region radius $\delta > 0$.
2: Choose $\lambda > -\lambda_n$.
3: **loop**
4: Compute the Cholesky factorization $LL^T = B + \lambda I$. L is lower triangular matrix
5: Solve $LL^T s = -g$. → solve $L^T s$
6: Solve $Lw = s$. → solve $Lw = s$
7: Set $\lambda \leftarrow \lambda + \Delta\lambda$ where

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

8: **end loop**

- Should always include a reasonable stopping criteria.
- Each iteration requires 1 factorization and 3 triangular-system solves.
- Choosing an initial $\lambda > -\lambda_n$ is not trivial.
- Method needs to be safeguarded to ensure convergence. See Chapter 7 of "Trust-Region Methods" by Conn, Gould and Toint for detailed discussion and analysis.

The Newton correction $\Delta\lambda$ at λ is given by

$$\Delta\lambda = -\phi(\lambda)/\phi'(\lambda).$$

If we differentiate

$$\phi(\lambda) = \frac{1}{\|s(\lambda)\|_2} - \frac{1}{\delta} = \frac{1}{\left[s(\lambda)^T s(\lambda)\right]^{\frac{1}{2}}} - \frac{1}{\delta}$$

we obtain

$$\phi'(\lambda) = -\frac{s(\lambda)^T \nabla s(\lambda)}{\left(s(\lambda)^T s(\lambda)\right)^{3/2}} = -\frac{s(\lambda)^T \nabla s(\lambda)}{\|s(\lambda)\|_2^3}.$$

Differentiating the defining equation $(B + \lambda I)s(\lambda) = -g$ with respect to λ yields

$$(B + \lambda I)\nabla s(\lambda) + s(\lambda) = 0 \quad \Longrightarrow \quad \nabla s(\lambda) = -(B + \lambda I)^{-1} s(\lambda).$$

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

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

is required to compute $\phi'(\lambda)$. Given the Cholesky factorization

$$B + \lambda I = L(\lambda)L(\lambda)^T$$

we can see that

$$\begin{aligned} s(\lambda)^T (B + \lambda I)^{-1} s(\lambda) &= s(\lambda)^T L(\lambda)^{-T} L(\lambda)^{-1} s(\lambda) \\ &= \left[L(\lambda)^{-1} s(\lambda)\right]^T \left[L(\lambda)^{-1} s(\lambda)\right] = \|w(\lambda)\|_2^2 \end{aligned}$$

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

Complexity results for Trust Region Subproblem

The problem

$$\underset{s \in \mathbb{R}^n}{\text{minimize}} \;\; m(s) \equiv f + s^T g + \tfrac{1}{2} s^T B s \;\; \text{subject to} \;\; \|s\|_2 \leq \delta$$

can be solved in polynomial time. More precisely, in the late 1980s and early 1990s, several researchers (Karmarkar, Ye, Vavasis, Zippel) showed the problem can be solved with at most $O\left(n^3 \log\left(\frac{1}{\epsilon}\right)\right)$ arithmetic operations. Ye (1992) improved this to $O\left(n^3 \log\log\left(\frac{1}{\epsilon}\right)\right)$ arithmetic operations, using interior point methods.

More generally, the problem

$$\underset{s \in \mathbb{R}^n}{\text{minimize}} \;\; \tfrac{1}{2} s^T Q_1 s + s^T b_1 \;\; \text{subject to} \;\; \tfrac{1}{2} s^T Q_2 s + s^T b_2 \leq \gamma$$

can be solved in polynomial time, i.e., $poly\left(n, \log\left(\frac{1}{\epsilon}\right)\right)$. Uses a concept called the **S-Lemma** and **Semidefinite Optimization/Programming (SDP)**.

Computing an exact solution to the trust-region subproblem is an iterative process

- each iteration requires a matrix factorization
- each iteration requires three triangular solves
- relatively expensive

Is it worth it?

- even an exact solution may be rejected if the trust-region radius is too large
- the steepest descent direction (very cheap to compute) makes good progress during early iterations
- recall why we used inexact searches in linesearch methods
- factorizations may not even be possible for very large problems
- we desire methods that find reasonable approximate solutions to the trust-region subproblem at a modest cost

We want to use the linear conjugate gradient (CG) method as the basis of an algorithm for computing an approximate solution p^{CG} to

$$\underset{p\in\mathbb{R}^n}{\text{minimize}} \;\; m(p) = f + g^T p + \tfrac{1}{2} p^T B p \quad \text{subject to} \quad \|p\|_2 \le \delta \tag{16}$$

- H may be indefinite (even singular) and we must ensure that

$$\Delta m(p^{CG}) \ge \Delta m(s^C)$$

where s^C is the Cauchy point for problem (16)

Theorem 2.9 (CG increasing norm property)

Suppose that the conjugate gradient (CG) method is applied to minimize $m(p)$ starting from $p_0 = 0$, and that the ith B-conjugate direction s_i generated by the CG algorithm satisfies $s_i^T B s_i > 0$ for all $0 \le i \le k$. Then the conjugate gradient iterates p_j satisfy

$$\|p_j\|_2 < \|p_{j+1}\|_2$$

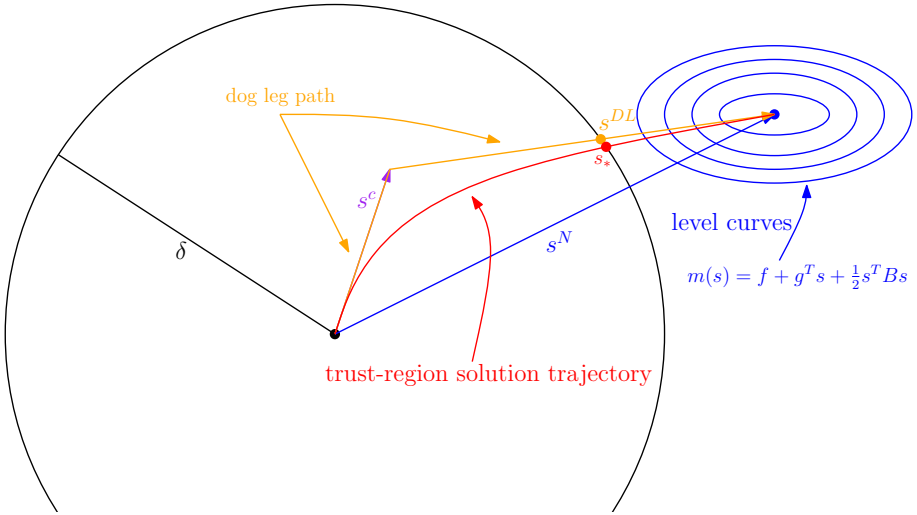
for all $0 \le j \le k-1$.

Basic idea: approximate the trust-region solution trajectory by a piecewise linear path

- define the dogleg step s^{DL} as the minimizer of $m(s)$ along the path

$$s(\alpha) = \begin{cases} \alpha s^C & \text{for } 0 \le \alpha \le 1 \\ s^C + (\alpha - 1)(s^N - s^C) & 1 \le \alpha \le 2 \end{cases}$$

  where s^C is the Cauchy point and s^N is the unconstrained Newton step
- it follows that $m(s^{DL}) \le m(s^C)$ so that $\Delta m(s^{DL}) \ge \Delta m(s^C)$, which ensures that the trust-region method is globally convergent
- care must be taken when B is singular



Algorithm 4 Steihaug Method based on linear CG

Input symmetric matrix $B \in \mathbb{R}^{n\times n}$ and vector g.
Choose stopping tolerance $\tau_{\text{stop}} > 0$.
Set $p_0 = 0$, $r_0 \leftarrow g$, $s_0 \leftarrow -g$, and $k \leftarrow 0$.
while $\|r_k\| > \tau_{\text{stop}}\|r_0\|$ **do**
  if $s_k^T B s_k > 0$ **then**
    Set $\alpha_k \leftarrow (r_k^T r_k)/(s_k^T B s_k)$.
  else
    Set $p_{k+1} \leftarrow p_k + \tau s_k$, where τ is the positive root of $\|p_k + \tau s_k\|_2 = \delta$.
    return $p^{CG} := p_{k+1}$
  end if
  if $\|p_k + \alpha_k s_k\|_2 < \delta$ **then**
    Set $p_{k+1} \leftarrow p_k + \alpha_k s_k$.
  else
    Set $p_{k+1} \leftarrow p_k + \tau s_k$, where τ is the positive root of $\|p_k + \tau s_k\|_2 = \delta$.
    return $p^{CG} := p_{k+1}$
  end if
  Set $r_{k+1} \leftarrow r_k + \alpha_k B s_k$.
  Set $\beta_{k+1} \leftarrow (r_{k+1}^T r_{k+1})/(r_k^T r_k)$.
  Set $s_{k+1} \leftarrow -r_{k+1} + \beta_{k+1} s_k$.
  Set $k \leftarrow k+1$.
end while
return $p^{CG} := p_k$

Crucially, it is easily seen that

$$m(p^{CG}) \leq m(s^{C}) \quad \text{and} \quad \|p^{CG}\|_2 \leq \delta$$

which implies that

$$\Delta m(p^{CG}) \geq \Delta m(s^{C})$$

so that (under standard assumptions) the trust-region algorithm with trial step p^{CG} converges to a first-order solution.

Fine, but how good is the trial step p_{CG}?

Theorem 2.10 (How good is the Steihaug truncated CG step?)

Suppose that the Steihaug truncated conjugate gradient method given by Algorithm 4 is applied to minimize $m(s)$ and that B is positive definite. Then the CG step p^{CG} and the global minimizer s_* satisfy the bound

$$\Delta m(p^{CG}) \geq \tfrac{1}{2}\Delta m(s_*)$$

In the non-convex case, the approximate solution p^{CG} may be very poor

- can we do better?

Outline of the generalized Lanczos trust-region (GLTR) method

The problem of interest

Find an approximate solution of the trust-region subproblem

$$\underset{s\in\mathbb{R}^n}{\text{minimize}} \; m(s) \quad \text{subject to} \quad \|s\|_2 \leq \delta$$

- solve a sequence of problems of the form

$$s^i = \underset{s\in\mathcal{Q}^i}{\text{argmin}} \; m(s) \quad \text{subject to} \quad \|s\|_2 \leq \delta$$

  where $\mathcal{Q}^i = \{Q^i y : y \in \mathbb{R}^i\}$
- the solution s^i satisfies $s^i = Q^i y^i$, where

$$y^i = \underset{y\in\mathbb{R}^i}{\text{argmin}} \; \|g\|_2 e_1^T y + \tfrac{1}{2} y^T T^i y \quad \text{subject to} \quad \|y\|_2 \leq \delta$$

- T^i is tri-diagonal and has a very sparse factorization so that the global minimizer can be computed efficiently using the earlier secular equation approach
- can exploit the structure and use the solution from one subproblem to initialize the solution process for solving the next subproblem
- until the trust-region boundary is reached, it is conjugate gradients.....but changes after the trust-region boundary is reached

Idea: minimize $m(s)$ subject to the trust-region constraint over a sequence of expanding subspaces (analogous to linear CG method)

How do we solve the problem over a subspace?

- instead of the basis of B-conjugate directions as in CG, use an equivalent Lanczos orthonormal basis
- Gram-Schmidt applied to CG (Krylov) basis

$$S^i \overset{\text{def}}{=} [s_0 \; s_1 \; \ldots \; s_{i-1}] \quad \text{for } i = 1, 2, \ldots n$$

  gives orthonormal basis

$$Q^i \overset{\text{def}}{=} [q_0 \; q_1 \; \ldots \; q_{i-1}] \quad \text{for } i = 1, 2, \ldots n$$

  such that $\mathbf{span}(S^i) = \mathbf{span}(Q^i)$
- Subspace $\mathcal{Q}^i = \{s : s = Q^i y \;\text{ for some }\; y \in \mathbb{R}^i\}$
- basis matrix Q^i satisfies

$$Q^{i\,T} Q^i = I \quad \text{and} \quad Q^{i\,T} B Q^i = T^i$$

  where T^i is tridiagonal and $Q^{i\,T} g = \|g\|_2 e_1$
- the basis matrix Q^i is trivial (cheap) to generate from the CG basis matrix S^i