Lecture 25 (Nov/30)

Scribe? Please fill the course evaluations.

Last time

- Trust region methods
- Characterization of subproblem
- How about other norms?

Today

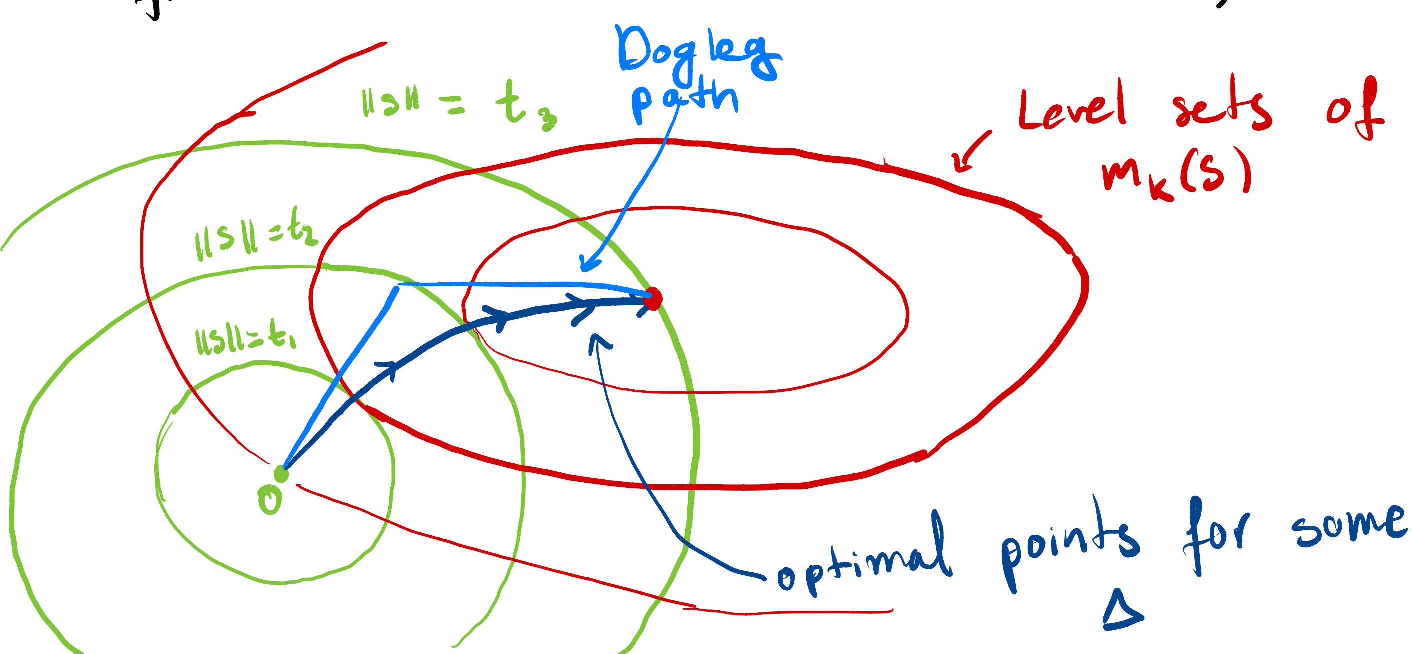
- Heuristics to solve the subproblem.
- Descent
- Full method
- Guarantees.

Recall the Trust Region Steps involve solving the nonconvex minimization problems

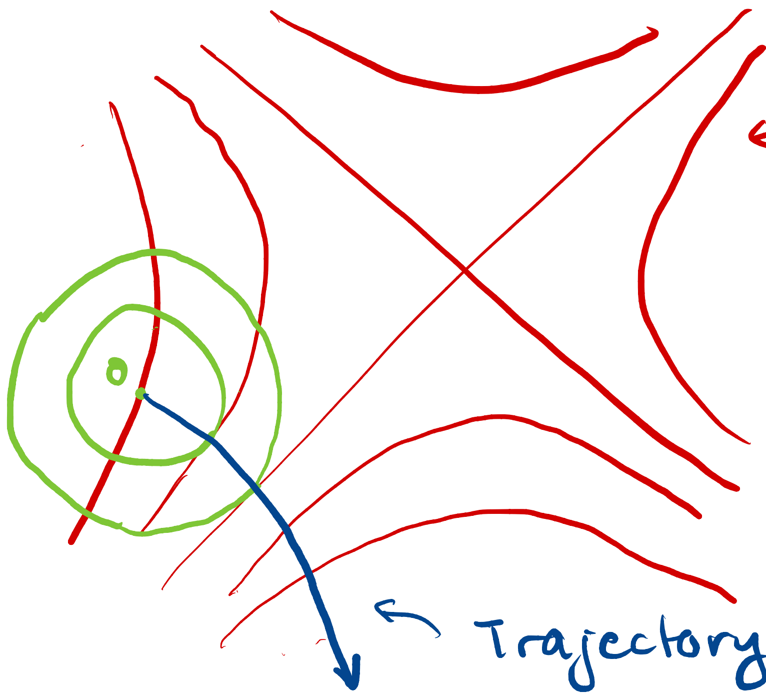
$$(\star)\quad s_k = \operatorname{argmin}\ m_k(s) = f(x_k) + \nabla f(x_k)^T s + \frac{s^T B_k s}{2} \quad \text{s.t. } \|s\|_2 \le \Delta_k$$

How does the solution change as we vary Δ_k?

If $m_k(s)$ is convex $(B_k \succ 0)$, then



If $m_k(s)$ is nonconvex



When $B_k \succ 0$, then we can approximate the trajectory via the so-called dogleg path:

$$s^{DL}(\tau) = \begin{cases} \tau s^{GD} & \text{if } \tau \in [0,1], \\ s^{GD} + (\tau - 1)(s^N - s^{GD}) & \text{if } \tau \in [1,2], \end{cases}$$

where

$$s^{GD} = -\left(\frac{\|g\|^2}{g^T B g}\right) g \quad \text{and} \quad s^N = -B^{-1} g.$$

This doesn't work when B_k is indefinite.

Then we can select

$$\operatorname*{argmin}_{\tau} \; m_k(s^{DL}(\tau))$$

$$\text{s.t.} \quad \underbrace{\| s^{DL}(\tau) \|}_{\text{and this increases.}} \le \Delta_k$$

One can show that this decreases with τ

Another "dogleg" heuristic considers

$$s_k = \operatorname*{argmin} \; m_k(s)$$
$$\text{s.t.} \quad \|s\| \le \Delta$$
$$s \in \operatorname{span}\{g_k, B_k^{-1} g_k\}.$$

These are only approximations!

There are other Linear Algebra approaches we don't cover:

- Gould et al. '99 "Solving the trust-region-subproblem using the Lanczos Method."
- Adachi et al. '17 "Solving the trust region-subproblem by a generalized eigenvalue problem."

Descent

Decrease is only guaranteed if our

approximation is good, i.e., Δ is small. Define the model objective decrease as

$$\Delta m_k(s) = m_k(0) - m_k(s) \quad (>0)$$

and function decrease as

$$\Delta f_k(s) = f(x_k) - f(x_k + s) \quad (\gtrless 0)$$

Lemma If f has L-Lips gradient, then for all $\|s\|_2 \le \Delta_k$

$$|\Delta f_k(s) - \Delta m_k(s)| \le \frac{1}{2}(L + \|B_k\|)\Delta_k^2.$$

If f has Q-Lipschitz Hessians, then

$$|\Delta f_k(s) - \Delta m_k(s)| \le \frac{Q}{6}\Delta_k^3 + \frac{\|B_k - \nabla^2 f(x_k)\|}{2}\Delta_k^2$$

Proof: Expanding

$$|\Delta f_k(s) - \Delta m_k(s)| = \Big| f(x_k + s) - (f(x_k) + \nabla f(x_k)^T s) - \frac{1}{2} s^T B_k s \Big|$$

Triangle inequality

$$\le |f(x_k + s) - (f(x_k) + \nabla f(x_k)^T s| + \frac{1}{2}|s^T B_k s|$$

Taylor

$$\le \frac{L}{2}\|s\|^2 + \frac{1}{2}\|B_k\| \|s\|^2.$$

$$\leq \left(\frac{L}{2} + \|B_k\|\right) \Delta_k^2 .$$

Similarly

$$|\Delta f_k(s) - \Delta m_k(s)| \leq \left| f(x_k+s) - \left(f(x_k) + \nabla f(x_k)^T s + \frac{1}{2} s^T \nabla^2 f(x_k) s\right)\right|$$

$$+ \frac{1}{2} \left| s^T (\nabla^2 f(x_k) - B_k) s \right|$$

Taylor

$$\leq \frac{a}{6} \|s\|^3 + \frac{1}{2} \|\nabla^2 f(x_k) - B_k\| \|s\|^2 .$$

$\square$

Next we show that the "Cauchy point" ensure some amount of model descent.

Define

$$s^c = \underset{\text{s.t.}}{\operatorname{argmin}} \quad \left\{ f + g^T s + \frac{1}{2} s^T B s \right\}$$

$$\|s\| \leq \Delta$$

$$s \in \operatorname{span}\{g\}$$

$$= \begin{cases} -\dfrac{\Delta}{\|g\|} g & \text{if } \Delta g^T B g \leq \|g\|^3 \\[2ex] -\left(\dfrac{\|g\|^2}{g^T B g}\right) g & \text{otherwise.} \end{cases}$$

Lemma The Cauchy point has

$$\Delta m_k(s^c) \geq \frac{1}{2}\|\nabla f(x_k)\| \min\left[\frac{\|\nabla f(x_k)\|}{\|B_k\|}, \Delta_k\right].$$

$\dashv$

Proof: Consider two cases

1. If $\Delta_k g_k^T B_k g_k \leq \|g_k\|^2$ with $g_k = \nabla f(x_k)$

$$\Rightarrow \Delta m_k(s^c) = \Delta\|g_k\| - \frac{1}{2}\Delta^2 \frac{g^T B_k g_k}{\|g_k\|^2}$$

$$\geq \Delta\|g_k\| - \frac{1}{2}\Delta\|g_k\|$$

$$\geq \frac{1}{2}\Delta\|g_k\|.$$

2. Otherwise

$$\Delta m_k(s^c) = \frac{\|g_k\|^4}{g^T B_k g} - \frac{1}{2}\frac{\|g_k\|^4}{g^T B_k g}$$

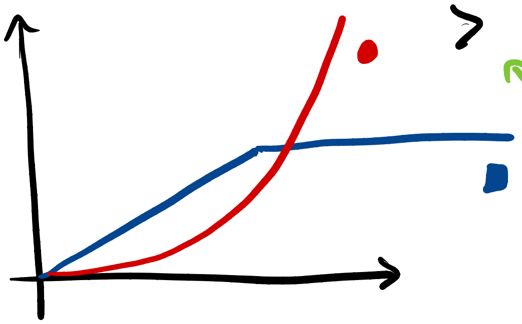
$$\geq \frac{1}{2}\frac{\|g_k\|^4}{g^T B_k g}$$

$$\geq \frac{1}{2}\frac{\|g_k\|^2}{\|B_k\|} \quad \text{since } g_k^T B_k g \leq \|g_k\|^2\|B_k\|.$$

Taken together these yield

$$\Delta f_k(s) \geq \Delta m_k(s) - \frac{1}{2}(L + \|B_k\|)\Delta_k^2$$

$$\geq \frac{1}{2}\|\nabla f(x_k)\| \min\left\{\frac{\|\nabla f\|}{\|B_k\|}, \Delta_k\right\} - \frac{1}{2}(L + \|B_k\|)\Delta_k^2$$



$> 0.$

For small enough Δ_k.

Full Trust Region Method

We measure how much we trust a step s by

$$\rho_k(s) = \frac{\Delta f_k(s)}{\Delta m_k(s)}$$

Goes to 1 as $\Delta_k \to 0$ $\Updownarrow$ $s \to 0$

If $\rho_k(s)$ is close to 1, this is a great step!

If $\rho_k(s)$ near zero or negative, this is a bad step!

Pick thresholds $0 < \eta_s \leq \eta_{vs} \leq 1$, x_0, Δ_0

Iterate:

- Find s_k minimizing $m_k(s)$ s.t. $\|s\| \leq \Delta_k$

At least as well as Cauchy.

▷ If $\rho_k(s_k) \geq \eta_{vs}$:

$$x_{k+1} = x_k + s_k$$
$$\Delta_{k+1} = 2\Delta_k$$

Else if $\rho_k(s_k) \geq \eta_s$:

$$x_k = x_k + s_k$$

Else:

$$x_{k+1} = x_k$$
$$\Delta_{k+1} = \Delta_k / 2.$$

Convergence Guarantees

Theorem (Global convergence, 2018, Curtis, Lubberts, Robinson)

Let $f: \mathbb{R}^d \to \mathbb{R}$ be C^2 function with a-Lipschitz Hessians, with $\inf f > -\infty$. Then, the trust region method (with additional checks) finds an ε-stationary point after $O(\varepsilon^{-2})$ iterations.

Theorem (4.9, Nocedal & Wright)

If $\|B_k - \nabla^2 f(x_k)\| \to 0$. Then, locally the method displays superlinear convergence.