Chapter 4 Trust Region methods

Back to unconstrained optimization. $\min_{x \in \mathbb{R}^n} f(x)$

Line search

$$p_k = -B_k^{-1} \nabla f_k \quad \text{(choose descent direction)}$$

find α_k $\quad f(x_k + \alpha_k p_k) - f(x_k) \leq \sigma \alpha_k \nabla f(x_k)^T p_k \quad , \quad \sigma \in (0,1)$

(SDC).

Trust region : another globalization procedure (guarantees convergence of QN methods from points far away from local min)

Find step direction and size at the same time by solving the trust region subproblem

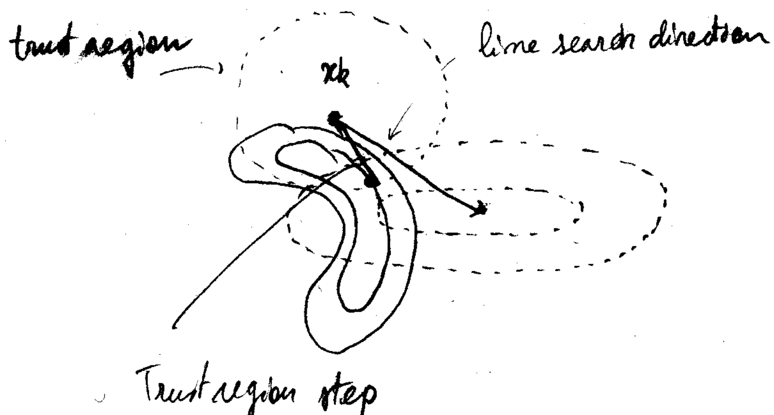
$$(TR) \begin{cases} \min & m_k(p) \\ \text{s.t.} & \|p\|_2 \leq \Delta_k \end{cases}$$

$\longrightarrow$ model usually quadratic:
$m_k(p) = f_k + \nabla f_k^T p + \frac{1}{2} p^T A_k p$

$\hookrightarrow$ describes ball around x_k in which model is believed to be a good approx of $f(x_k + p)$

$\hookrightarrow$ another norm such as l_1-norm could be used, but we focus on 2-norm.

Here is a situation where TR gives a better step than line search:



Trust Region framework

Monitor agreement between model m_k and objective function.

Since we care about reducing f we look at:

$$\boxed{\rho_k = \frac{ared_k}{pred_k} = \frac{f(x_k+p_k) - f(x_k)}{m_k(p_k) - m_k(0)} = \frac{\text{actual reduction}}{\text{predicted reduction}}}$$

solves problem on TR. ∈ trust region

since $m_k(p_k) < m_k(0)$:

if $\rho_k < 0 \Rightarrow f(x_k+p_k) - f(x_k) \Rightarrow$ step must be rejected, model is not that good (Δ_k must be reduced)

if $\rho_k \sim 1$: good agreement keep going — increase Δ_k

if $0 < \rho_k \ll 1$ — keep same Δ_k.

step acceptance threshold

Given $\Delta_{max} > 0$, $\Delta_0 \in (0, \Delta_{max})$ and $\eta \in [0, 1/4)$

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

solve (TR) for p_k (approximate solution is OK)

$\rho_k = \frac{ared_k}{pred_k}$

TR adjustment:

if $\rho_k < \frac{1}{4}$ then $\Delta_{k+1} = \frac{1}{4}\Delta_k$ (reduce TR)

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

$\Delta_{k+1} = \min(2\Delta_k, \Delta_{max})$ (increase TR up to Δ_{max})

else

$\Delta_{k+1} = \Delta_k$ (keep same TR)

if $\rho_k > \eta$

$x_{k+1} = x_k + p_k$; accept step

else

$x_{k+1} = x_k$; reject step

<u>Note</u>: TR radius is increased only if TR constraint is active ($\|p_k\| = \Delta_k$).
If $\|p_k\| < \Delta_k$ then TR is not interfering with finding step

Solution of TR subproblem

(TR) $\min f_k + \nabla f_k^T p + \frac{1}{2} p^T H_k p$
$\|p\|_2 \leq \Delta_k$

<u>Theorem</u> p^* is a global solution of (TR) iff
$\|p^*\| \leq \Delta_k$ and $\exists\, \lambda^* \geq 0$ s.t.

$$(H_k + \lambda I)\, p^* = -\nabla f_k \qquad (1)$$

$$\lambda(\Delta_k - \|p^*\|) = 0 \qquad (2)$$

$$H_k + \lambda I \succeq 0 \quad (\text{pos semi def}) \qquad (3)$$

<u>proof</u>: 2nd order optimality conditions (KKT) for constrained opt..

$$L(p, \lambda) = f_k + \nabla f_k^T p + \frac{1}{2} p^T H_k p - \lambda(\Delta - \|p\|^2)$$

(1) $\Leftrightarrow$ $\nabla_p L(p^*, \lambda^*) = 0$
(2) $\Leftrightarrow$ complementarity. either $\lambda^* = 0$ or $\Delta_k = \|p^*\|$.
(3) $\Leftrightarrow$ $\nabla^2_{pp} L(p^*, \lambda^*)$ is pos semi def

Approximate solutions to TR subproblem

We do not need to solve (TR) exactly. To get convergence, approximate sol to (TR) must satisfy

$$m_k(p_k) - m_k(0) \leq \underset{(0,1)}{\underset{\in}{\sigma}} \left(m_k(p^{CP}) - m_k(0)\right)$$

↑ the reduction in the model must be
at least a fraction of decrease obtained by <u>Cauchy point</u>

Cauchy point: go in direction of negative gradient and take into account TR boundary.

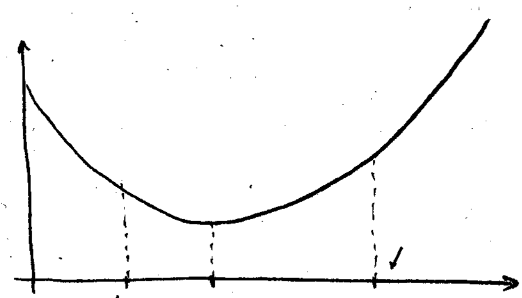
$p^{cp} = -t^{cp} \nabla f(x_k)$ where

$$t^{cp} = \begin{cases} \dfrac{\|\nabla f_k\|_2^2}{\nabla f_k^T H_k \nabla f_k} & \text{if } \nabla f_k^T H_k \nabla f_k > 0 \text{ and } \dfrac{\|\nabla f_k\|_2^3}{\nabla f_k^T H_k \nabla f_k} \le \Delta_k \\[2ex] \dfrac{\Delta_c}{\|\nabla f_k\|} & \text{otherwise} \end{cases}$$

Easy to see as a one dimensional optimization problem:

$$m_k(-t \nabla f_k) = f_k - t\|\nabla f_k\|^2 + \frac{t^2}{2} \nabla f_k^T H_k \nabla f_k \quad \text{(quad in } t\text{)}$$

- $\nabla f_k^T H_k \nabla f_k > 0$

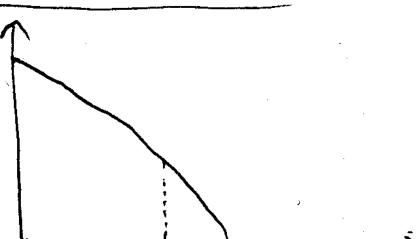


$\frac{\Delta_c}{\|\nabla f_k\|}$ t_{cp} $\frac{\Delta_c}{\|\nabla f_k\|}$ t

when $\geq$

when $\dfrac{\|\nabla f_k\|_2^3}{\nabla f_k^T H_k \nabla f_k} \le \Delta_k$

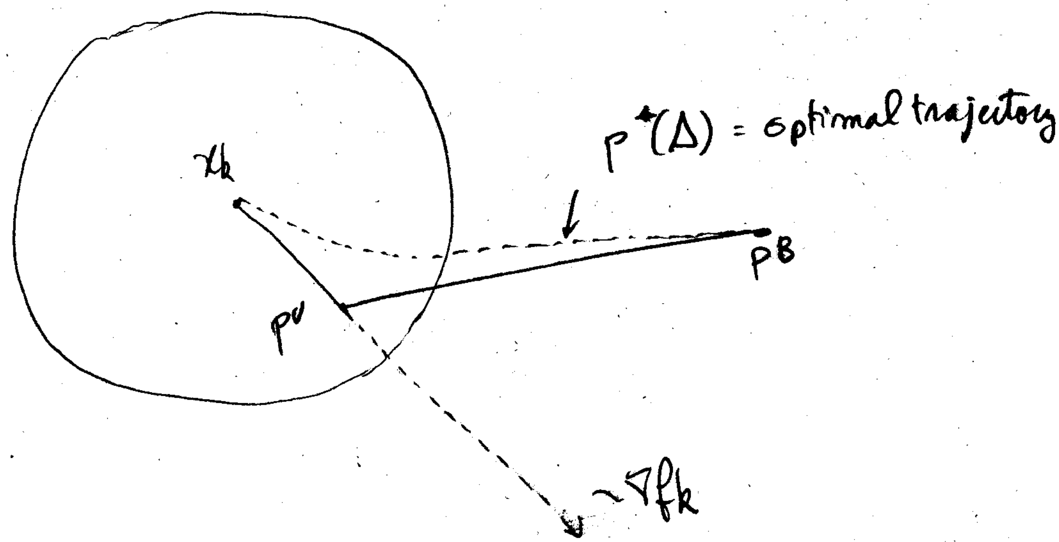
- $\nabla f_k^T H_k \nabla f_k \le 0$



$\frac{\Delta_c}{\|\nabla f_k\|}$

Dogleg method to improve on Cauchy point

Let $\vec{p}(\Delta)$ be solution to TR problem as a function of TR radius Δ.



p^U = unconstrained min along gradient direction.

$$= -\frac{\|\nabla f_k\|_2^2}{\nabla f_k^T H_k \nabla f_k} \nabla f_k$$

p^B = full step (i.e. we minimize m_k without worrying about TR)

$$= -H_k^{-1} \nabla f_k$$

form dogleg path:

$$\tilde{p}(\tau) = \begin{cases} \tau p^U & \tau \in [0,1] \\ p^U + (\tau - 1)(p^B - p^U) & \tau \in [1,2] \end{cases}$$

Idea: instead of solving (TR), minimize function along path.

it can be shown:

$\|\tilde{p}(\tau)\|$ increases with τ.

$m(\tilde{p}(\tau))$ decreases with τ

if $\|p^B\| \geq \Delta$ then dogleg path intersects TR in one single point which can be found by solving:

$$\|p^U + (\tau-1)(p^B - p^U)\|^2 = \Delta^2$$

(quadratic)

if $\|p^B\| \leq \Delta$ choose p^B.

Convergence results

$$m_k(0) - m_k(p_k^{CP}) \geq \frac{1}{2}\|\nabla f_k\| \min\left(\Delta_k, \frac{\|\nabla f_k\|}{\|H_k\|}\right) \quad (*)$$

same if dogleg is used since

$$m_k(p^{dogleg}) \leq m(p_k^{CP})$$

we allow to reject steps

Theorem Let $\eta \in (0, \frac{1}{4})$ in TR algo. Suppose $\|H_k\| \leq \beta$, f is bdd below and Lipschitz continuously diff'ble and that all iterates satisfy $(*)$ and

$\|p_k\| \leq \gamma \Delta_k$, for some $\gamma \geq 1$. (we allow steps to go beyond TR but not by much)

Then $\lim_{k \to \infty} \|\nabla f_k\| = 0$.