Chapter 4 Trust Region methods

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

Line search
\[ p_k = -B_k^{-1} \nabla f_k \quad \text{(some descent direction)} \]
find \( \alpha_k \) such that \( f(x_k + \alpha_k p_k) - f(x_k) \leq \alpha_k \nabla f(x_k)^T p_k \), \( \alpha_k \in (0, 1) \)
(SDC).

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

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

\[
\begin{align*}
\min_{p, \|p\|_2 \leq \Delta_k} & \quad m_k(p) \\
\text{s.t.} & \quad \|p\|_2 \leq \Delta_k
\end{align*}
\]

\( m_k(p) = f_k + \nabla f_k^T p + \frac{1}{2} p^T H_k p \)

\( \rightarrow \) model usually quadratic:

\( m_k(p) = f_k + \nabla f_k^T p + \frac{1}{2} p^T H_k p \)

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

\( \rightarrow \) another norm such as \( L_1 \)-norm could be used, but we focus on \( L_2 \)-norm.

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

trust region \( x_k \)
line search direction

Trust region step
Trust Region framework

Monitor agreement between model and objective function.

Since we are about reducing \( f \) we look at:

\[
\frac{\text{actual reduction}}{\text{predicted reduction}} = \frac{f(x_k + p_k) - f(x_k)}{m_k(p_k) - m_k(0)}
\]

\( \text{where problem } x_k \in \text{trust region} \)

Since \( m_k(p_k) < m_k(0) \):

- \( S_k = \frac{\text{actual reduction}}{\text{predicted reduction}} \)
  - \( f(x_k + p_k) - f(x_k) \Rightarrow \) step must be rejected, model is not that good (\( \Delta_k \) must be reduced)
  - Increase \( \Delta_k \)
  - Keep same \( \Delta_k \)
  - Step acceptance threshold

- \( 0 < S_k < 1 \)
  - good agreement, keep going

- \( S_k > 1 \)
  - \( \Delta_k \) is too large, reduce it

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

\( k = 0, 1, 2, \ldots \)

Solve (TR) for \( p_k \) (approximate solution is OK)

\( S_k = \frac{\text{actual reduction}}{\text{predicted reduction}} \)

\[
\begin{cases} 
\text{if } S_k < \frac{1}{4} & \text{then } \Delta_{k+1} = \frac{1}{4} \Delta_k \quad \text{(reduce TR)} \\
\text{else if } S_k > \frac{3}{4} & \text{and } \|p_k\| = \Delta_k \text{ then } \\
& \Delta_{k+1} = \min(2\Delta_k, \Delta_{\text{max}}) \quad \text{(increase TR up to } \Delta_{\text{max}}) \\
\text{else} \\
& \Delta_{k+1} = \Delta_k \quad \text{(keep same TR)} \\
\end{cases}
\]

- \( S_k > \eta \)
  - \( x_{k+1} = x_k + p_k \), accept step

- \( S_k > \eta \)
  - \( x_{k+1} = x_k \), reject step
Note: 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**

$$(TR) \quad \min \ p_k + \nabla f_k^T p + \frac{1}{2} p^T H_k p$$

$\|p\| \leq \Delta_k$

**Theorem**

$p^*$ is a global solution of (TR) iff

$\|p^*\| < \Delta_k$ and $\exists \; \lambda^* > 0$ s.t.

$$\begin{align}
(H_k + \lambda I) p^* &= - \nabla f_k \\
\lambda (\Delta_k - \|p^*\|) &= 0 \\
H_k + \lambda I &\succeq 0 \quad \text{(pos semi-def)}
\end{align}$$

**Proof:** 2nd order optimality conditions (KKT) for constrained opt.

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

(1)$\iff$ $\nabla_L L(p^*, \lambda^*) = 0$

(2) complementarity, either $\lambda^* > 0$ or $\Delta_k = \|p^*\|$

(3) $\nabla^2_L L(p^*, \lambda^*)$ 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

$$\Delta_k (p_k) - \Delta_k (0) \leq \epsilon (\Delta_k (p^{ep}) - \Delta_k (0))$$

At least a fraction of decrease obtained by Cauchy point.
Canedy point: go in direction of negative gradient and take into account TR boundary.

\[ p^c = -t^c \nabla f(0_k) \]

where

\[ t^c = \begin{cases} \frac{\|\nabla f_k\|^2}{\nabla f_k^T H_k \nabla f_k} & \text{if } \nabla f_k^T H_k \nabla f_k > 0 \text{ and } \frac{\|\nabla f_k\|^2}{\nabla f_k^T H_k \nabla f_k} \leq \Delta_k \\ \frac{\Delta c}{\|\nabla f_k\|} & \text{otherwise} \end{cases} \]

Easy to see as a one-dimensional optimization problem:

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

\[ \nabla f_k^T H_k \nabla f_k > 0 \]

\[ \text{when } \frac{\|\nabla f_k\|^2}{\nabla f_k^T H_k \nabla f_k} \leq \Delta_k \]

\[ \nabla f_k^T H_k \nabla f_k < 0 \]

\[ \text{when } \frac{\|\nabla f_k\|^2}{\nabla f_k^T H_k \nabla f_k} \geq \Delta_k \]
Dogleg method to improve an Cauchy point

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

$\hat{p}^*(\Delta) =$ optimal trajectory

$\hat{p}^u =$ unconstrained min along gradient direction.

\[ \hat{p}^u = \frac{-\nabla f_k}{\nabla f_k^T H_k \nabla f_k} \nabla f_k \]

$\hat{p}^B =$ full step (i.e. we minimize rank without worrying about TR)

\[ \hat{p}^B = -H_k^{-1} \nabla f_k \]

Dogleg path:

\[ \hat{p}(z) = \begin{cases} z \hat{p}^u & z \in [0, 1] \\ \hat{p}^u + (z-1)(\hat{p}^B - \hat{p}^u) & z \in [1, 2] \end{cases} \]

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

It can be shown:

- $\| \hat{p}(z) \|$ increases with $z$.
- $m(\hat{p}(z))$ decreases with $z$.
if \( \| p^k \| \geq \Delta \) then degley path intersects TR in one single point which can be found by solving:
\[
\| p^u + (r-1)(p^0 - p^u) \| = \Delta^2
\]
(quadatic)

if \( \| p^k \| \leq \Delta \) choose \( p^0 \).

**Convergence results**

\[
m_k(0) - m_k(p^0_{\text{degly}}) \geq \frac{1}{2} \| Df_k \| \min (\Delta k, \frac{\| Df_k \|}{\| h_k \|}) \tag{*}
\]

some of degley is used since
\[
m_k(p_{\text{degly}}) \leq m(p_{\text{degly}})
\]

we allow to reject steps

**Theorem** Let \( \eta \in (0, \frac{1}{4}) \) in TR algo. Suppose \( \| h_k \| \leq \beta \), \( f \) is odd below and Lipschitz continuously diffble and that all iterates satisfy \((*)\) and
\[
\| p^k \| < \gamma \Delta k \quad \text{for some } \gamma > 1. \quad \text{(we allow steps to go beyond TR but not by much)}
\]

Then lim \( \| Df_k \| = 0 \).

\( k \to \infty \)