CS5321 Numerical Optimization

04 Trust Region Methods

Trust region method

1. Solve model problem m_k. Let p_k be the solution.
2. Evaluate p_k and update the trust region.

- The quadratic model will be used.

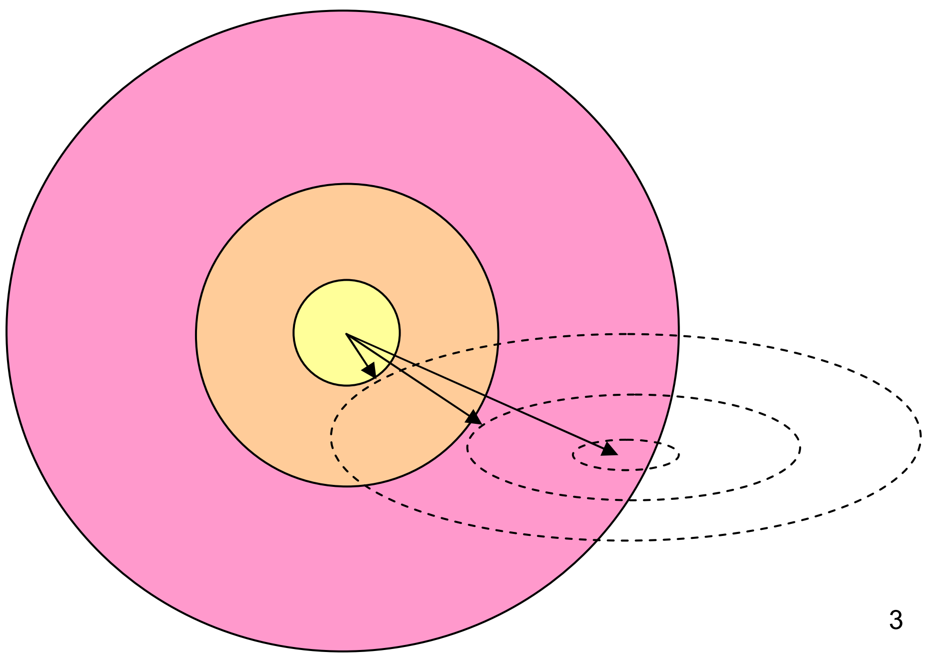
$$\min_{\|p\| \leq \Delta_k} m_k(p) = f_k + g_k^T p + \frac{1}{2} p^T B_k p$$

  - $f_k = f(x_k)$, B_k is the Hessian, and g_k is the gradient.
  - Δ_k is the trust region radius

Solve model problem m_k

$$\min_{\|p\| \le \Delta_k} m_k(p) = f_k + g_k^T p + \frac{1}{2} p^T B_k p$$

- The problem is a constrained optimization
- If $\|B^{-1}g\| \le \Delta$ and B is positive definite, $p = -B^{-1}g$.
- Otherwise, the direction varies for different Δ.



Optimal conditions

- p^* is the optimal solution if and only if it satisfies

$$(B+\lambda I)\, p^* = -g$$

$$\lambda(\Delta - \|p^*\|) = 0$$

$$B+\lambda I \text{ is positive semidefinite}$$

  - $\lambda \geq 0$. is called Largrangian modifier (chap 12)

- Assume $\|B^{-1}g\| \geq \Delta_k$. for $\lambda \geq 0$. Define $\phi(\lambda) = \|-(B+\lambda I)^{-1} g\| - \Delta$ and solve $\phi(\lambda)=0$
  - This is a univariable nonlinear equation. (chap 11)

Approximation methods

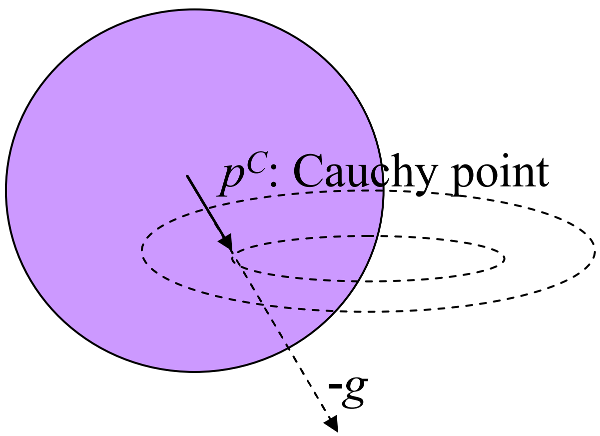
- The problem $\phi(\lambda)=\|-(B+\lambda I)^{-1}g\|-\Delta=0$ is difficult to formulate and solve
  - Can be used when the number of variables is small.
- Four approximate methods
  1. Cauchy point
  2. The dogleg method
  3. Two-dimensional subspace minimization
  4. Steihaug's algorithm (chap 7)

1. Cauchy point

- The steepest descent method + line search
- The solution is $p_k = -\tau_k \frac{\Delta_k}{\|g_k\|} g_k$ (Cauchy point)

$$\tau_k = \begin{cases} 1 & \text{if } g_k^T B_k g_k \leq 0 \\ \min(\|g_k\|^3/(\Delta_k g_k^T B_k g_k, 1)) & \text{otherwise.} \end{cases}$$

  - Slow convergence
  - Easy to compute
  - Use as a reference direction
  - More on this in chap 5



2. Dogleg method

- Require B_k to be positive definite.
- Use $p(\tau)$ to approximate the optimal trajectory

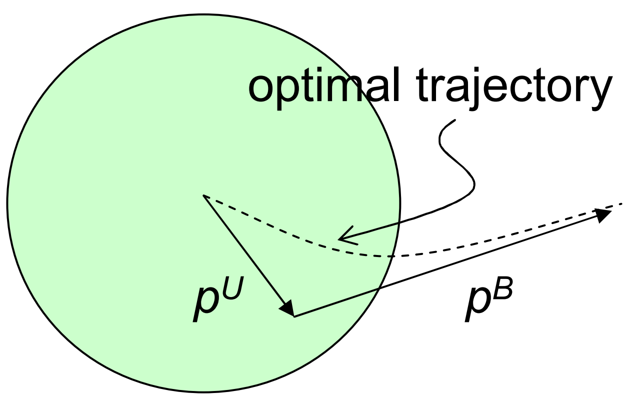
$$p(\tau) = \begin{cases} \tau p^U & 0 \le \tau \le 1 \\ p^U + (\tau - 1)(p^B - p^U) & 1 < \tau \le 2 \end{cases}$$

  - p^{U} is the Cauchy point

$$p^U = -\frac{g^T g}{g^T B g} g$$

  - p^{B} is the Newton's direction

$$p^B = -B^{-1} g$$



3. 2-dim subspace minimization

- Use the linear combination of g and $B^{-1}g$

$$\min_{\|p\|\le\Delta_k} m_k(p) = f_k + g_k^T p + \frac{1}{2}p^T B_k p \text{ s.t. } p \in \text{span}\{g, B^{-1}g\}$$

- Matrix B can be indefinite.
  - Find α such that $(B + \alpha I)$ is positive definite
  - If $\| (B + \alpha I)^{-1} g \| \le \Delta_k$, $p = -(B + \alpha I)^{-1} g + v$ such that $v^T (B + \alpha I)^{-1} g \le 0$.
  - Otherwise, $p \in \text{span}\{g, (B + \alpha I)^{-1} g\}$

Evaluation function

- Solution p_k is evaluated by

$$\rho_k = \frac{f(x_k) - f(x_k + p_k)}{m_k(0) - m_k(p_k)}$$

  - Numerator $f(x_k) - f(x_k + p_k)$ is the *actual* reduction
  - Denominator $m_k(0) - m(p_k)$ is the *predicted* reduction
  - If ρ_k is close to 1, m_k is a good model. In this case, if $\|p_k\| = \Delta_k$, increase Δ_k.
  - If ρ_k is close to 0 or negative, shrink the range of the trust region Δ_k.

Scaling

- Poor scaled problems are sensitive to certain directions
  - Ex: $f(x, y) = 10^9 x^2 + y^2$.
- Solution 1: if knowing the scale of final solution, one can rescale the variables
- Solution 2: The trust region can be elliptical.

$$\min_{\|Dp\| \leq \Delta_k} m_k(p) = f_k + g_k^T p + \frac{1}{2} p^T B_k p$$

  - D is a diagonal matrix

Global convergence

- Theorem 4.5

  Suppose $\|B_k\|$ is bounded, and f is bounded below on the level set $S=\{x| f(x) \leq f(x_0)\}$ and Lipschitz continuously differentiable in the neighborhood of S. If

$$m_k(0) - m_k(p_k) \geq c_1 \|g_k\| \min\left(\Delta_k, \frac{\|g_k\|}{\|B_k\|}\right)$$

  and $\|p_k\| \leq \gamma\Delta_k$ for some $c_1 \in (0,1]$ and $\gamma \geq 1$. Then

$$\lim_{k\to\infty} \inf \|g_k\| = 0$$