CS5321 Numerical Optimization

02 Fundamental of Unconstrained Optimization

What is a solution?

- A point x^* is a global minimizer if $f(x^*) \leq f(x)$ for all x.
- A point x^* is a local minimizer if there is a neighborhood $\mathcal{N}$ of x^* such that $f(x^*) \leq f(x)$ for all $x \in \mathcal{N}$.

Necessary conditions

- If x^* is a local minimizer of f and $\nabla^2 f$ exists and is continuously differentiable in an open neighborhood of x^*, then $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ is positive semidefinite
  - x^* is called a *stationary point* if $\nabla f(x^*) = 0$.

Convexity

- When f is convex, any local minimizer x^* is a global minimizer of f.
- In addition, if f is differentiable, then any stationary point x^* is a global minimizer of f.

Two strategies

1. Line search method (chap 3)
   a) Given a point x_k, find a descent direction p_k.
   b) Find the step length α_k to minimize $f(x_k+\alpha_k p_k)$
   c) Next point $x_{k+1} = x_k + \alpha_k p_k$.
2. Trust region method (chap 4)
   a) For a function f, construct a model function m_k.
   b) Define a trust region $\mathcal{R}(x_k)$ inside which $f \approx m_k$.
   c) Solve the minimization problem: $\min_p m_k(x_k+p)$ where p lies inside $\mathcal{R}(x_k)$

Methods to find descent directions for line search

- Method 1: the steepest descent method
  - First order approximation
  - Linear convergence (global convergence)
- Method 2: the Newton's method
  - Second order approximation
  - Quadratic convergence (local convergence)
- Method 3: the conjugate gradient method (chap 5)

The steepest descent method

- The steepest descent method (first order approx)
  - From the Taylor's expansion,

$$f(x_k + p) \approx f(x_k) + p^T \nabla f(x_k)$$

  - To minimize $p^T \nabla f(x_k)$, such that $\|p\| = 1$

$$p = -\nabla f(x_k) / \|\nabla f(x_k)\|$$

    - Because $p^T \nabla f(x_k) = \|p\| \|\nabla f(x_k)\| \cos\theta$
      Minimization is at $\cos\theta = -1$

The Newton's method

- The Newton's method (second order approx)
  - Assume the Hessian is positive definite
  - From the Taylor's expansion,

    $$f(x_k + p) \approx f(x_k) + p^T \nabla f(x_k) + \frac{1}{2} p^T \nabla^2 f(x_k) p$$

  - The minimization of $f(x_k+p)$ is at $\nabla f(x_k + p) = 0$
    - Note p is the variable. $f(x_k)$, $\nabla f(x_k)$, $\nabla^2 f(x_k)$ are constants.

      $$\nabla f(x_k + p) \approx \nabla f(x_k) + \nabla^2 f(x_k) p = 0$$

    - Newton's direction is $p^N = -(\nabla^2 f(x_k))^{-1} \nabla f(x_k)$

Substitute $p^N = -(\nabla^2 f(x_k))^{-1}\nabla f(x_k)$ to

$$f(x_k + p) \approx f(x_k) + p^T \nabla f(x_k) + \tfrac{1}{2} p^T \nabla^2 f(x_k) p$$

$$\begin{aligned}
f(x_k + p^N) &\approx f(x_k) + (p^N)^T \nabla f(x_k) + \frac{1}{2}(p^N)^T \nabla^2 f(x_k) p^N \\
&= f(x_k) - \nabla f(x_k)^T (\nabla^2 f(x_k))^{-1} \nabla f(x_k) + \\
&\quad \frac{1}{2} \nabla f(x_k)^T (\nabla^2 f(x_k))^{-1} \nabla f(x_k) \\
&= f(x_k) - \frac{1}{2} \nabla f(x_k)^T (\nabla^2 f(x_k))^{-1} \nabla f(x_k) \\
&\leq f(x_k)
\end{aligned}$$

Variations of Newton's method

- The Hessian matrix may be indefinite, ill-conditioned, or even singular.
  - Modified Newton's methods. (chap 3)
- The computation of Hessian is expensive
  - Quasi-Newton methods (chap 6)
- The inverse of Hessian is expensive
  - Inexact Newton's method (chap 7)

The conjugate gradient method

- The current search direction p_k is a linear combination of previous search direction p_{k-1} and current gradient

$$p_k = -\nabla f(x_k) + \beta_k p_{k-1}$$

- Scalar β_k is given such that p_k and p_{k-1} are *conjugate*.

Models for trust-region method

- Linear model:

$$f(x_k + p) \approx f(x_k) + p^T \nabla f(x_k)$$

- Quadratic model:

$$f(x_k + p) \approx f(x_k) + p^T \nabla f(x_k) + \frac{1}{2} p^T \nabla^2 f(x_k) p$$

  - All variations of Newton's method can be applied.

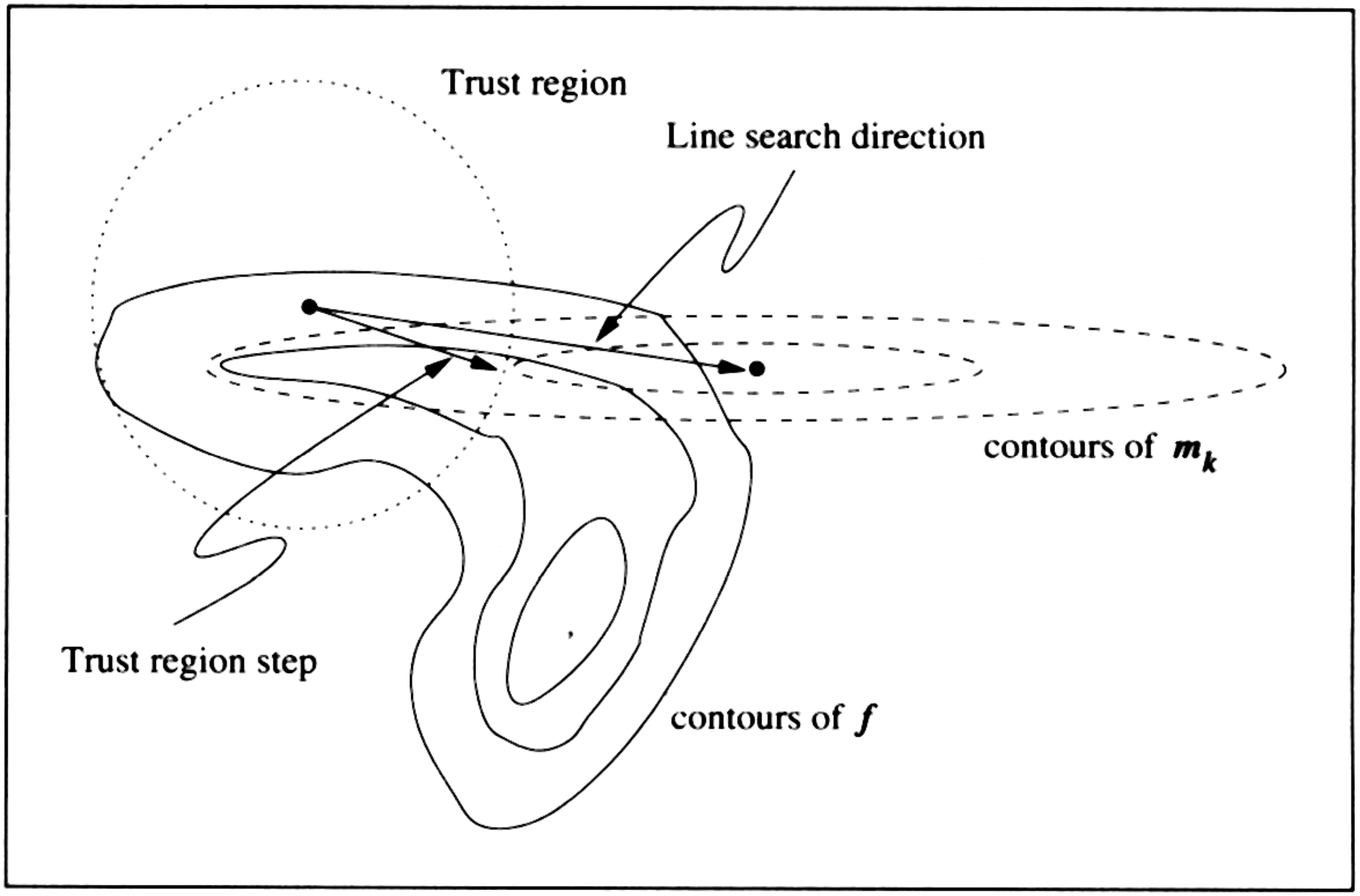

Figure 4.1 Trust-region and line search steps.