02 - Fundamentals of Unconstrained Optimization

September 5, 2020

0.1 MATH692 - Numerical Optimization

1 02 Fundamentals of Unconstrained Optimization

1.0.1 The problem is

$$\min_{x \in \mathbb{R}^n} f(x)$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a real-valued smooth function defined on the Euclidean space $\mathbb{R}^n$.

1.1 What is a solution?

Global Solution A point x^* is a **global minimizer** if $f(x^*) \leq f(x)$ for all x

Local Solution A point x^* is a **weak (strong) local minimizer** if there is a neighborhood $\mathcal{N}$ of x^* such that

$$f(x^*) \leq (<) f(x) \quad \text{for all } x \in \mathcal{N}(\backslash\{x^*\})$$

1.1.1 Theorem (Taylor's Theorem).

Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable and that $p \in \mathbb{R}^n$. Then we have that

$$f(x+p) = f(x) + \nabla f(x+tp)^T p, \text{ for some } t \in (0,1).$$

Moreover, if f is twice continuously differentiable, we have that

$$\nabla f(x+p) = \nabla f(x) + \int_0^1 \nabla^2 f(x+tp) p dt,$$

and that

$$f(x+p) = f(x) + \nabla f(x)^T p + \frac{1}{2} p^T \nabla^2 f(x+tp) p, \text{ for some } t \in (0,1).$$

Remark For convex functions every **local minimizer** is also a **global minimizer**. (why?)

1.1.2 Theorem (First-Order Necessary Conditions).

If x^* is a local minimizer and f is continuously differentiable in an open neighborhood of x^*, then x^* is a **stationary point**, i.e.

$$\nabla f(x^*) = 0.$$

1.1.3 Remark

When f is convex and differentiable, then any stationary point x^* is a global minimizer of f.

1.1.4 Theorem (Second-Order Necessary Conditions).

If x^* is a local minimizer of f and $\nabla^2 f$ exists and is continuous in an open neighborhood of x^*, then

$$\nabla f(x^*) = 0 \text{ and } \nabla^2 f(x^*) \text{ is positive semidefinite.}$$

1.1.5 Theorem (Second-Order Sufficient Conditions).

Suppose that $\nabla^2 f$ is continuous in an open neighborhood of x^* and that $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ is positive definite. Then x^* is a strict local minimizer of f .

1.2 Convex sets and Convex functions

Convex Sets: A set $S \in \mathbb{R}^n$ is a **convex set** if the straight line segment connecting any two points in S lies entirely inside S. i.e. for any two points $x \in S$ and $y \in S$, we have

$$\alpha x + (1 - \alpha)y \in S \quad \text{for all } \alpha \in [0, 1].$$

Convex Functions: The function f is a **convex function** if its domain S is a convex set and if for any two points x and y in S, the following property is satisfied:

$$f(x + (1 -)y) \; f(x) + (1 -)f(y), \text{ for all } \; [0, 1].$$

1.2.1 Theorem.

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

1.2.2 Algorithms

- For unconstrained optimization of smooth functions (chapters: 3,4,5,6,7).
- Beginning at x_0, **optimization algorithms** generate a sequence of iterates $\{x_k\}$ that terminate when either no more progress can be made or when it seems that a solution point has been approximated with sufficient accuracy.
- An algorithm finds a new iterate x_{k+1} with a lower function value than x_k.

1.2.3 Two strategies

1. Line search methods (chapter 3)
2. Trust Region methods (chapter 4)

1.2.4 Line Search

1. Given a point x_k, find a **descent direction** p_k.
2. Find the step length α_k to minimize $f(x_k + \alpha_k p_k)$. i.e. we solve
$$\min_{\alpha > 0} f(x_k + \alpha p_k)$$
3. Next point $x_{k+1} = x_k + \alpha_k p_k$.

1.2.5 Trust Region

1. For a function f, construct a **model function** m_k that approximates f locally.
2. Define a trust region $\mathcal{R}(x_k)$ inside which $f \approx m_k$.
3. Solve the minimization problem:

$$\min_p m_k(x_k + p) \text{ where } p \text{ lies inside } \mathcal{R}(x_k)$$

1.2.6 Methods to find descent directions for line search

-
-
-

1.2.7 Method 1: the steepest descent method $p_k = -\nabla f(x_k)$

- First order approximation
- Linear convergence (global convergence)

1.2.8 Method 2: the Newton's method

- Second order approximation: $p_k^N = -\left[\nabla^2 f(x_k)\right]^{-1} \nabla f(x_k)$ where the Hessian $\nabla^2 f(x_k)$ is assumed to be positive definite.
- Newton direction is a descent direction.
- Step size is $\alpha = 1$.
- Quadratic convergence (local convergence)

Problems with the Hessian $\nabla^2 f(x_k)$

- The Hessian matrix may be indefinite, illconditioned, 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)

1.2.9 Method 3: the conjugate gradient method (chap 5)

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

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

- Scalar k is given such that p_k and p_{k-1} are conjugate (definition later).

1.2.10 Models for Trust-Region Methods

- Linear model:

  $$m_k := f_k + p^T \nabla f_k$$

  and hence we have to solve

  $$\min_p f_k + p^T \ f_k \text{ subject to } \|p\|^2 \ \Delta_k.$$

- Quadratic model:

  $$m_k := f_k + p^T \nabla f_k + \frac{1}{2} p^T \nabla^2 f_k p$$

  and hence we have to solve

  $$\min_p m_k \text{ subject to } \|p\|^2 \ \Delta_k.$$

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

[ ]:

```

```