无约束优化

2013.12.30

- 1.无约束优化问题
- 2.下降法
- 3.梯度下降法
- 4.最速下降法
- 5.牛顿法

1.无约束优化问题

- 问题描述

$$\text{minimize} \quad f(x)$$

其中，$f(x)$ 为凸函数，且二次可微。

- 两种解法
  - **求解梯度方程：** $\nabla f(x^*) = 0$
  - **迭代逼近：** $f(x^{(k)}) \to p^*$

1.无约束优化问题

- 强凸性

  函数 $f(x)$在 S上具有强凸性,若存在*m*>0,使得

$$\nabla^2 f(x) \succeq mI \qquad \text{for all } x \in S$$

- 强凸性的作用：确定停止准则

2.下降法

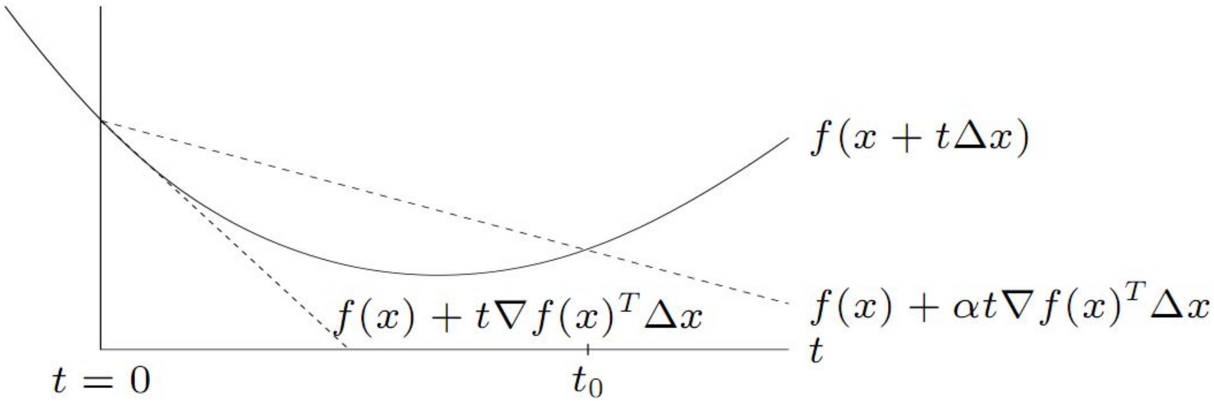
- 基本原理

  迭代 $x^{(k+1)} = x^{(k)} + t\Delta x$ ，满足 $f(x^{(k+1)}) < f(x^{(k)})$
  - 其中，Δx为下降方向，t为步长因子；
  - 对于凸函数 $f(x)$ 当Δx满足 $\nabla f(x)^T \Delta x < 0$时，存在某个$t$，使得 $f(x^{(k+1)}) < f(x^{(k)})$。
- 算法
  - 给出初始点
  - 循环迭代，直到满足停止准则
    - **计算下降方向：** Δx
    - **搜索步长因子：** $t > 0$
    - 更新　： $x := x + t\Delta x$

2.下降法

- 步长因子搜索方法
  - 精确一维搜索

    $$t = \mathrm{argmin}_{t>0} f(x + t\Delta x)$$
  - 回溯一维搜索
    - **初始化：令** $t = 1$，**给定参数** $\alpha \in (0, 0.5), \beta \in (0,1)$
    - **循环迭代** $t = \beta t$ ，**直到** $f(x+t\Delta x) < f(x) + \alpha t \nabla f(x)^T \Delta x$



3.梯度下降法

- 下降方向：$\Delta x = -\nabla f(x)$
- 终止条件：$\|\nabla f(x)\|_2 \leq \epsilon$
- 算法简单，但收敛速度较慢。

4.最速下降法

- 归一化最速下降方向：

$$\Delta x_{\mathrm{nsd}} = \operatorname{argmin}\{\nabla f(x)^T v \mid \|v\| = 1\}$$

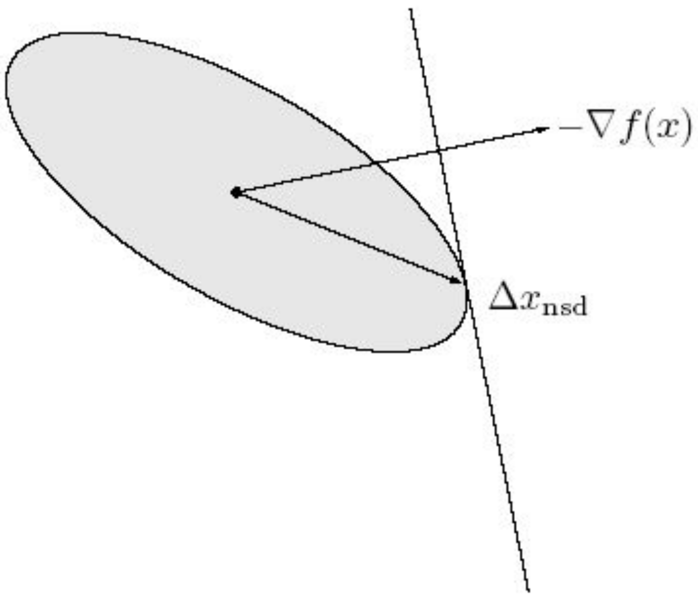
- 非归一化最速下降方向：

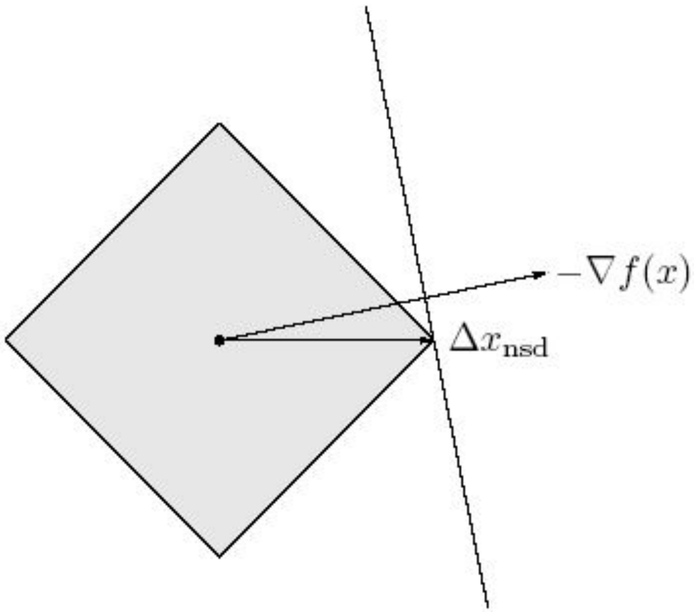
$$\Delta x_{\mathrm{sd}} = \|\nabla f(x)\|_* \Delta x_{\mathrm{nsd}}$$

- 欧式范数：$\Delta x_{\mathrm{sd}} = -\nabla f(x)$
- 二次范数$\|x\|_P = (x^T P x)^{1/2}$：$\Delta x_{\mathrm{sd}} = -P^{-1}\nabla f(x)$
- ℓ_1 范数：$\Delta x_{\mathrm{sd}} = -(\partial f(x)/\partial x_i)e_i$

4.最速下降法

- 规范化的最速下降方向是 $\|.\|$ 的单位球体中在 $-\nabla f(x)$ 的方向上投影最长的方向





5. 牛顿法

- 下降方向：$\Delta x_{\mathrm{nt}} = -\nabla^2 f(x)^{-1} \nabla f(x)$
  - 牛顿方向为二次范数 $\|x\|_{\nabla^2 f(x)} = (x^T \nabla^2 f(x) x)^{1/2}$上的最速下降方向。
- 牛顿减量(用于设计停止准则)：

$$\lambda(x) = \left(\nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x)\right)^{1/2}$$

- 下降方向及牛顿减量都仿射不变。

5. 牛顿法

- 算法
  - 给出初始点
  - 循环迭代
    - 计算下降方向和减量：

$$\Delta x_{\rm nt} := -\nabla^2 f(x)^{-1}\nabla f(x)$$

$$\lambda^2 := \nabla f(x)^T\nabla^2 f(x)^{-1}\nabla f(x)$$

    - 若$\lambda^2/2 \le \epsilon$ 则终止退出；
    - 一维线性搜索，计算步长因子 t
    - 更新： $x := x + t\Delta x$

5. 牛顿法

- 牛顿法的优点：
  - 一般情况下收敛快，二次收敛。
  - 仿射不变性，对坐标选择或目标函数的下水平集不敏感。
  - 和问题规模有很好的比例关系，高维及低维中问题的性能相似，迭代次数增加不大。
  - 不依赖算法参数的选择。
- 牛顿法的缺点：
  - 计算及存储Hessian矩阵及计算下降方向需要较高的成本

谢谢！