无约束可微问题的最优性理论:引言

无约束可微优化问题通常表示为如下形式：

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

其中f 是连续可微函数.

- 给定一个点$\bar{x}$, 我们想要知道这个点是否是函数f 的一个局部极小解或者全局极小解.
- 如果从定义出发, 需要对其邻域内的所有点进行判断, 这不可行.
- 因此, 需要一个更简单的方式来验证一个点是否为极小值点.我们称其为最优性条件, 它主要包含一阶最优性条件和二阶最优性条件.

一阶必要条件:下降方向

定义 (下降方向)

对于可微函数f 和点$x \in \mathbb{R}^n$, 如果存在向量d 满足

$$\nabla f(x)^{\mathrm{T}} d < 0,$$

那么称d 为f 在点x 处的一个下降方向.

- 一阶最优性条件是利用梯度(一阶)信息来判断给定点的最优性.
- 由下降方向的定义, 容易验证: 如果f 在点x 处存在一个下降方向d, 那么对于任意的$T > 0$, 存在$t \in (0, T]$, 使得

$$f(x + td) < f(x).$$

因此, 在局部最优点处不能有下降方向.

一阶必要条件

定理 (一阶必要条件)

假设f 在全空间$\mathbb{R}^n$ 可微.如果x^* 是一个局部极小点, 那么

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

证明:任取$v \in \mathbb{R}^n$, 考虑f 在点$x = x^*$ 处的泰勒展开

$$f(x^* + tv) = f(x^*) + tv^\mathrm{T}\nabla f(x^*) + o(t),$$

整理得

$$\frac{f(x^* + tv) - f(x^*)}{t} = v^\mathrm{T}\nabla f(x^*) + o(1).$$

根据x^* 的最优性, 在上式中分别对t 取点0 处的左, 右极限可知

$$\lim_{t\to 0^+} \frac{f(x^* + tv) - f(x^*)}{t} = v^\mathrm{T}\nabla f(x^*) \geq 0,$$
$$\lim_{t\to 0^-} \frac{f(x^* + tv) - f(x^*)}{t} = v^\mathrm{T}\nabla f(x^*) \leq 0,$$

即对任意的v 有$v^\mathrm{T}\nabla f(x^*) = 0$, 由$v$ 的任意性知$\nabla f(x^*) = 0$.

一阶必要性条件:注记

- 对于$f(x) = x^2, x \in \mathbb{R}$, 我们知道满足$f'(x) = 0$ 的点为$x^* = 0$, 并且其也是全局最优解.
- 对于$f(x) = x^3, x \in \mathbb{R}$, 满足$f'(x) = 0$ 的点为$x^* = 0$, 但其不是一个局部最优解.
- 称满足$\nabla f(x) = 0$ 的点x 为f 的稳定点(有时也称为驻点或临界点).
- 除了一阶必要条件, 还需要对函数加一些额外的限制条件, 才能保证最优解的充分性.

二阶最优性条件

- 在没有额外假设时，如果一阶必要条件满足，我们仍然不能确定当前点是否是一个局部极小点.
- 假设f 在点x 的一个开邻域内是二阶连续可微的.类似于一阶必要条件的推导，可以借助当前点处的二阶泰勒展开来逼近该函数在该点附近的取值情况，从而来判断最优性.
- 在点x 附近我们考虑泰勒展开

$$f(x+d)=f(x)+\nabla f(x)^{\mathrm{T}}d+\frac{1}{2}d^{\mathrm{T}}\nabla^2 f(x)d+o(\|d\|^2).$$

- 当一阶必要条件满足时，$\nabla f(x)=0$，那么上面的展开式简化为

$$f(x+d)=f(x)+\frac{1}{2}d^{\mathrm{T}}\nabla^2 f(x)d+o(\|d\|^2).$$

由此，我们可以导出二阶最优性条件.

二阶最优性条件

定理 (二阶最优性条件)

必要条件:若x^* 是f 的一个局部极小点, 则$\nabla f(x^*) = 0, \nabla^2 f(x^*) \succeq 0$.

充分条件:若$\nabla f(x^*) = 0, \nabla^2 f(x^*) \succ 0$, 则$x^*$ 是f 的一个局部极小点.

- **必要性**:若$\nabla^2 f(x^*)$ 有负的特征值$\lambda_- < 0$, 设$\nabla^2 f(x^*)d = \lambda_- d$, 则

$$\frac{f(x^*+d) - f(x^*)}{\|d\|^2} = \frac{1}{2}\frac{d^{\mathrm{T}}}{\|d\|}\nabla^2 f(x^*)\frac{d}{\|d\|} + o(1) = \frac{1}{2}\lambda_- + o(1).$$

  当$\|d\|$ 充分小时,$f(x^*+d) < f(x^*)$, 这和点x^* 的最优性矛盾.

- **充分性**:由$\nabla f(x^*) = 0$ 时的二阶展开,

$$\frac{f(x^*+d) - f(x^*)}{\|d\|^2} = \frac{d^{\mathrm{T}}\nabla^2 f(x^*)d + o(\|d\|^2)}{\|d\|^2} \geq \frac{1}{2}\lambda_{\min} + o(1).$$

  当$\|d\|$ 充分小时有$f(x^*+d) \geq f(x^*)$, 即二阶充分条件成立.

- 设点$\bar{x}$ 满足一阶最优性条件(即$\nabla f(\bar{x}) = 0$), 且该点处的海瑟矩阵$\nabla^2 f(\bar{x})$ 不是半正定的, 那么$\bar{x}$ 不是一个局部极小点.
- 进一步地, 如果海瑟矩阵$\nabla^2 f(\bar{x})$ 既有正特征值又有负特征值, 我们称稳定点$\bar{x}$ 为一个鞍点.
- 事实上, 记d_1, d_2 为其正负特征值对应的特征向量, 那么对于任意充分小的$t > 0$, 我们都有$f(\bar{x} + td_1) > f(\bar{x})$ 且$f(\bar{x} + td_2) < f(\bar{x})$.
- 注意, 二阶最优性条件给出的仍然是关于局部最优性的判断.对于给定点的全局最优性判断, 我们还需要借助实际问题的性质, 比如目标函数是凸的、非线性最小二乘问题中目标函数值为0 等.

- 线性最小二乘:

$$\min_{x\in\mathbb{R}^n} \quad f(x) \stackrel{\text{def}}{=\!=} \frac{1}{2}\|b-Ax\|_2^2,$$

其中$A\in\mathbb{R}^{m\times n}$, $b\in\mathbb{R}^m$.由f 可微且凸知

$$x^* \text{ 为一个全局最优解} \Leftrightarrow \nabla f(x^*) = A^\mathrm{T}(Ax^*-b) = 0.$$

- 非线性最小二乘(实数情形的相位恢复):

$$\min_{x\in\mathbb{R}^n} \quad f(x) \stackrel{\text{def}}{=\!=} \sum_{i=1}^m r_i^2(x),$$

其中$r_i(x) = (a_i^\mathrm{T}x)^2 - b_i^2, i = 1,2,\cdots,m$.

实数情形的相位恢复

$$\nabla f(x) = 2\sum_{i=1}^{m} r_i(x)\nabla r_i(x) = 4\sum_{i=1}^{m}((a_i^{\mathrm{T}}x)^2 - b_i^2)(a_i^{\mathrm{T}}x)a_i,$$

$$\nabla^2 f(x) = \sum_{i=1}^{m}(12(a_i^{\mathrm{T}}x)^2 - 4b_i^2)a_ia_i^{\mathrm{T}}.$$

- 如果x^* 为局部最优解, 那么其满足一、二阶必要条件

$$\sum_{i=1}^{m}((a_i^{\mathrm{T}}x^*)^2 - b_i)(a_i^{\mathrm{T}}x^*)a_i = 0, \quad \sum_{i=1}^{m}(12(a_i^{\mathrm{T}}x^*)^2 - 4b_i^2)a_ia_i^{\mathrm{T}} \succeq 0.$$

- 如果一个点$x^\#$ 满足二阶充分条件

$$\sum_{i=1}^{m}((a_i^{\mathrm{T}}x^\#)^2 - b_i)(a_i^{\mathrm{T}}x^\#)a_i = 0, \quad \sum_{i=1}^{m}(12(a_i^{\mathrm{T}}x^\#)^2 - 4b_i^2)a_ia_i^{\mathrm{T}} \succ 0,$$

那么$x^\#$ 为局部最优解.