Lecture 1 Notes

2021 年 9 月 8 日

1 预备知识

1.1 向量的范数

定义. 若实值函数 $\|\cdot\|:\mathbb{R}^n\to\mathbb{R}$ 满足：

(1). 正定性：$\|\boldsymbol{x}\|\geq 0(\forall \boldsymbol{x}\in\mathbb{R}^n)$ 且 $\|\boldsymbol{x}\|=0$ 当且仅当 $\boldsymbol{x}=\mathbf{0}$

(2). 齐次性：$\|\alpha\boldsymbol{x}\|=|\alpha|\|\boldsymbol{x}\|,\forall\alpha\in\mathbb{R},\forall\boldsymbol{x}\in\mathbb{R}^n$

(3). 三角不等式：$\|\boldsymbol{x}+\boldsymbol{y}\|\leq\|\boldsymbol{x}\|+\|\boldsymbol{y}\|,\forall\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^n$

称 $\|\cdot\|$ 为向量范数.(这里 $\mathbb{R}^n$ 表示 n 维欧氏空间)

设 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^\top\in\mathbb{R}^n$，常用到的向量范数有：

$$\|\boldsymbol{x}\|_1=\sum_{i=1}^n|x_i|$$

$$\|\boldsymbol{x}\|_2=(\sum_{i=1}^n x_i^2)^{\frac{1}{2}}$$

$$\|\boldsymbol{x}\|_\infty=\max_i|x_i|$$

1.2 梯度，海瑟矩阵，泰勒展开

假设集合 $S\subseteq\mathbb{R}^n$，函数 f 是定义在 S 上的 n 元实值函数，给出以下的记号：

$f\in C(S)$ 表示 f 在任意一点 $\boldsymbol{x}\in S$ 上连续

$f\in C^1(S)$ 表示在任意一点 $\boldsymbol{x}\in S$ 都有 $\frac{\partial f}{\partial x_i},(\forall i=1,2,\cdots n)$ 存在且连续

$f\in C^2(S)$ 表示在任意一点 $\boldsymbol{x}\in S$ 都有 $\frac{\partial^2 f}{\partial x_i\partial x_j},(\forall i,j=1,2,\cdots n)$ 存在且连续

函数 f 在 $\boldsymbol{x}$ 处的梯度定义为 n 维列向量，记为 $\nabla f(\boldsymbol{x})$：

$$\nabla f(\boldsymbol{x})=[\frac{\partial f}{\partial x_1}(\boldsymbol{x}),\frac{\partial f}{\partial x_2}(\boldsymbol{x}),\cdots,\frac{\partial f}{\partial x_n}(\boldsymbol{x})]^\top$$

函数 f 在 $\boldsymbol{x}$ 处的海瑟矩阵定义为 $n\times n$ 维矩阵，记为 $\nabla^2 f(\boldsymbol{x})$，$\nabla^2 f(\boldsymbol{x})$ 的第 i 行第 j 列元素为：

$$[\nabla^2 f(\boldsymbol{x})]_{i,j}=\frac{\partial^2 f(\boldsymbol{x})}{\partial x_i\partial x_j},1\leq i,j\leq n$$

若 $f \in C^1(S)$ 则给定点 $\bar{\boldsymbol{x}} \in S$，$f(\boldsymbol{x})$ 在 $\bar{\boldsymbol{x}}$ 处的一阶泰勒展开式为

$$f(\boldsymbol{x}) = f(\bar{\boldsymbol{x}}) + \nabla f(\bar{\boldsymbol{x}})^\top (\boldsymbol{x} - \bar{\boldsymbol{x}}) + o(\|\boldsymbol{x} - \bar{\boldsymbol{x}}\|)$$

这里 $o(\|\boldsymbol{x} - \bar{\boldsymbol{x}}\|)$ 表示的是当 $\|\boldsymbol{x} - \bar{\boldsymbol{x}}\| \to 0$ 时的高阶无穷小

若 $f \in C^2(S)$ 则给定点 $\bar{\boldsymbol{x}} \in S$，$f(\boldsymbol{x})$ 在 $\bar{\boldsymbol{x}}$ 处的二阶泰勒展开式为

$$f(\boldsymbol{x}) = f(\bar{\boldsymbol{x}}) + \nabla f(\bar{\boldsymbol{x}})^\top (\boldsymbol{x} - \bar{\boldsymbol{x}}) + \frac{1}{2}(\boldsymbol{x} - \bar{\boldsymbol{x}})^\top \nabla^2 f(\bar{\boldsymbol{x}})(\boldsymbol{x} - \bar{\boldsymbol{x}}) + o(\|\boldsymbol{x} - \bar{\boldsymbol{x}}\|^2)$$

2 最优化问题的概念

最优化问题一般可以建立为如下的数学模型，假设 $S \subseteq \mathbb{R}^n$：

$$\begin{aligned} \min_{\boldsymbol{x} \in S} \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & f_j(\boldsymbol{x}) = 0, (j = 1, 2, \cdots, k) \\ & f_j(\boldsymbol{x}) \le 0, (j = k+1, k+2, \cdots, m) \end{aligned}$$

这里 $f, f_1, f_2, \cdots, f_m : \mathbb{R}^n \to \mathbb{R}$ 的函数，一般称 $f(\boldsymbol{x})$ 为目标函数

定义上述问题的可行域如下，即在 S 中满足优化问题的等式约束与不等式约束的点的全体：

$$Q = \{\boldsymbol{x} \in S \,|\, f_j(\boldsymbol{x}) = 0, (j = 1, 2, \cdots, k), f_j(\boldsymbol{x}) \le 0, (j = k+1, k+2, \cdots, m)\}$$

若 $Q \neq \emptyset$ 则称问题是可行的.

若存在 $\boldsymbol{x}$ 是 Q 中的内点，且满足 $f_j(\boldsymbol{x}) = 0, j = 1, 2, \cdots, k$ 以及 $f_j(\boldsymbol{x}) < 0, j = k+1, k+2, \cdots, m$ 则称问题是严格可行的.

定义. 问题的全局极小点：给定点 $\boldsymbol{x}^* \in Q$ 若对于所有的 $\boldsymbol{x} \in Q$ 都有 $f(\boldsymbol{x}^*) \le f(\boldsymbol{x})$

定义. 问题的局部极小点：给定点 $\boldsymbol{x}^* \in Q$ 若存在 $\delta > 0$ 满足任意的 $\boldsymbol{x} \in Q \cap B(\boldsymbol{x}^*, \delta)$ 都有 $f(\boldsymbol{x}^*) \le f(\boldsymbol{x})$, (这里的 $B(\boldsymbol{x}^*, \delta) = \{\boldsymbol{x} | \|\boldsymbol{x} - \boldsymbol{x}^*\| \le \delta\}$)

2.1 优化问题的分类

- 有约束的优化问题 v.s. 无约束的优化问题

  可行域 Q 是否等于 $\mathbb{R}^n$

- 连续的优化问题 v.s. 离散的优化问题

$$x_i \in \mathbb{Z} \Leftrightarrow \sin(\pi \cdot x_i) = 0$$

$$x_i \in \{0, 1\} \Leftrightarrow x_i \cdot (x_i - 1) = 0$$

- 线性规划 v.s. 非线性规划

  目标函数 $f(\boldsymbol{x}) = \boldsymbol{c}^\top \boldsymbol{x}$，等式约束与不等式约束中的 $f_j(\boldsymbol{x}) = a_j^\top \boldsymbol{x} + b_j$，此外 $S = \mathbb{R}^n$ 这样的问题是线性规划问题. 否则是非线性规划问题

  若 $S = \mathbb{R}^n$，非线性规划与线性规划问题之间，可以看到如果非线性函数有光滑性，通过泰勒展开非线性规划问题可以被近似为线性规划问题.

- 非光滑问题 v.s. 光滑问题

  所谓的光滑性是指约束函数，目标函数的偏导数是否连续可微，下面的例子说明非光滑的函数是可以通过光滑的函数近似的，例子中是考虑二维欧氏空间 $\mathbb{R}^2$ 上的函数 $h(\boldsymbol{x})$

$$h(\boldsymbol{x}) = \max\{x_1, x_2\} \approx \frac{x_1 + x_2 + \sqrt{(x_1 - x_2)^2 + \epsilon}}{2}, \epsilon \searrow 0^+$$

3 数值方法的概念

对于数值优化问题求解得到精确解一般是很困难的，比如对于一个光滑的函数 $f(\boldsymbol{x})$ 考虑无约束的优化问题 $\min f(\boldsymbol{x})$，对于该问题利用必要性条件等价于求解方程组 $\nabla f(\boldsymbol{x}) = \boldsymbol{0}$，如果 $\boldsymbol{x}$ 是 1000 维相当于求解 1000 维的方程组，很难给出精确解 (这只是无约束优化问题而且 f 是光滑函数的情形) 因此求解最优化问题数值方法是不可或缺的。

3.1 数值方法的要素

首先数值方法都是针对某一类特定的问题进行求解，没有一种数值方法能做到对于所有的问题都能求解而且表现都很好，因此对于一个问题首先要明确它是属于哪一类的优化问题比如线性规划还是非线性规划等等

对于一个优化问题 P，设它属于的问题类别是 $\mathcal{F}$，此时因为要使用数值方法求解问题必定需要问题的一些信息比如目标函数的值，约束函数的值，目标函数的梯度，约束函数的梯度等等，定义给出问题的信息的过程是 oracle.(简单来说就是给定点 $\boldsymbol{x}$ 通过 oracle 这个过程最后我们可以得到我们数值方法中需要的信息)

求解的步骤概述：

1. 给出初值点 x_0，令 $k = 0$，初始信息 $I_{-1} = \emptyset$.
2. 利用 oracle 给出当前点 $\boldsymbol{x}^k$ 处的信息记为 $\mathcal{O}(\boldsymbol{x}^k)$.
3. 更新迭代获得的总信息 $I_k = I_{k-1} \cup (\boldsymbol{x}^k, \mathcal{O}(\boldsymbol{x}^k))$.
4. 利用当前的信息 I_k 以及数值方法得到新的迭代点 $\boldsymbol{x}^{k+1}$。
5. 检验新的点 $\boldsymbol{x}^{k+1}$ 是否满足终止准则，若满足输出 $\boldsymbol{x}^{k+1}$ 作为数值解，若不满足则令 $k = k + 1$ 转到步 2.

3.2 数值方法的评估

对于一个类内的问题可能会有很多种方法求解，在这些数值方法中针对具体的问题会有不同的表现最直观的就是求解同一个问题时数值方法的运行时间如果是不同，肯定会倾向去选择耗时短的数值方法，因此对于一个算法应该要给出一些评判的标准。

定义. *解析复杂度：给定精度 ϵ(这里的这个常数是用在求解步骤中的终止准则的设计上) 此时给出 oracle 以及数值方法，则可以按照上述的求解步骤设计算法，在算法终止时调用 oracle 去获取信息的总次数称为算法的解析复杂度*

算术复杂度：给定精度 ϵ(这里的这个常数是用在求解步骤中的终止准则的设计上) 此时给出 oracle 以及数值方法，则可以按照上述的求解步骤设计算法，算法终止时基本运算的总次数 (也就是 oracle 操作需要的加减乘除次数和数值方法更新得到迭代点的加减乘除次数总和)