优化方法：前三周课后答疑

龚梓阳（助教）

日期：2021 年 9 月 25 日

1 基础知识

在此处，我将根据自己的认识和同学的提问，对部分学过以及部分可能尚未涉及的知识进行一个简单地梳理及介绍。

关于多变量函数 $f:\mathbb{R}^n\to\mathbb{R}$ 的相关定义，如可微、可导、连续，方向导数和泰勒公式等，可类似于单变量函数推广得到，具体说明可以参考“数学分析”（复旦欧阳光中）中第 13，14 章（多变量微积分学）。其余关于矩阵的知识，如正定、半正定的定义可参考“高等代数”（北大）第 5 章（二次型）。

1.1 范数

定义 1.1 (范数)**.** 设 X 是域 $\mathbb{K}$ （实数域或复数域）上的线性空间，函数 $\|\cdot\|:X\to\mathbb{R}$ 满足：

1. （正定性）$\forall x\in X, \|x\|\geq 0$；$\|x\|=0 \iff x=0$；
2. （齐次性）$\forall x\in X, \alpha\in\mathbb{K}, \|\alpha x\|=|\alpha|\cdot\|x\|$；
3. （次可加性）$\forall x,y\in X, \|x+y\|\leq\|x\|+\|y\|$。

则称 $\|\cdot\|$ 是 X 上的一个范数。

例 1.1 (常见范数)**.** 常见范数有：

1. 空间 $\mathbb{R}$：$\forall x\in\mathbb{X}$
$$\|x\|:=|x| \tag{1}$$
即范数 $\|x\|$ 为 x 的绝对值。
2. 空间 $\mathbb{R}^n$：$\forall \mathbf{x}\in\mathbb{R}^n, \mathbf{x}=(x_1,x_2,\ldots,x_n)'$
   (a) 欧几里得范数：
$$\|\mathbf{x}\|_2:=\sqrt{x_1^2+\cdots+x_n^2} \tag{2}$$
   (b) l_p 范数 $(p\geq 1)$：
$$\|\mathbf{x}\|_p:=\left(\sum_{i=1}^n |x_i|^p\right)^{1/p} \tag{3}$$
   若 $p=1$，则
$$\|\mathbf{x}\|_1:=\sum_{i=1}^n |x_i| \tag{4}$$
   若 $p=\infty$，则
$$\|\mathbf{x}\|_\infty:=\max_{i=1,\ldots,n}|x_i| \tag{5}$$

(c) l_0 范数：

$$\|\mathbf{x}\|_0 = |\{x_i : x_i \neq 0, i = 1, \ldots, n\}| \tag{6}$$

即向量 $\mathbf{x}$ 中分量 x_i 不为零的个数。

评论. 关于范数的由来，以及更详细的定义和性质，将在未来的“泛函分析”这门课程中详细进行学习。在本门课程中，所涉及的 $\|x-y\|$ 可以理解为两个数 x, y 之间距离远近的一种度量方式。

1.2 梯度向量与 Hessien 矩阵

定义 1.2 (梯度向量)**.** 对于多变量函数 $f : \mathbb{R}^n \to \mathbb{R}$，即 $f(\mathbf{x}) = f(x_1, x_2, \cdots, x_n)$，如果 f 在点 $\mathbf{x}_0$ 关于每一个变量 x_i 都有偏导数 $\frac{\partial f}{\partial x_i}(\mathbf{x}_0)$ 存在，则在点 $\mathbf{x}$ 上，这些偏导数定义了一个向量:

$$\nabla f(\mathbf{x}_0) = \left(\frac{\partial f}{\partial x_1}(\mathbf{x}_0), \ldots, \frac{\partial f}{\partial x_n}(\mathbf{x}_0)\right)' \tag{7}$$

该向量称为 f 在点 $\mathbf{x}_0$ 的梯度向量。

评论. 关于该定义，在“数学分析”（复旦欧阳光中）中第 14.6 节（方向导数与梯度）中有过简单介绍，各位同学可翻阅一下自己的教科书，或者可以借助搜索引擎对梯度有一个更清晰的认识。

定义 1.3 (Hessian 矩阵)**.** 对于多变量函数 $f : \mathbb{R}^n \to \mathbb{R}$, 如果 f 在点 $\mathbf{x}_0$ 的所有二阶偏导数都存在，那么函数 f 在点 $\mathbf{x}_0$ 的 Hessian 矩阵为

$$\boldsymbol{H}(\mathbf{x}_0) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(\mathbf{x}_0) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(\mathbf{x}_0) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(\mathbf{x}_0) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(\mathbf{x}_0) & \frac{\partial^2 f}{\partial x_2^2}(\mathbf{x}_0) & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(\mathbf{x}_0) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(\mathbf{x}_0) & \frac{\partial^2 f}{\partial x_n \partial x_2}(\mathbf{x}_0) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(\mathbf{x}_0) \end{bmatrix} \tag{8}$$

或使用下标记号表示为

$$\boldsymbol{H}_{ij}(\mathbf{x}_0) = \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}_0) \tag{9}$$

1.3 凸集与凸函数

以下内容摘抄自课程用书“数值最优化方法”附录 1，对于完成作业可能有帮助。

定义 1.4 (凸集)**.** 设集合 $C \subset \mathbb{R}^n$。若对 $\forall \mathbf{x}, \mathbf{y} \in C$，有

$$\theta\mathbf{x} + (1-\theta)\mathbf{y} \in C, \quad \theta \in [0, 1] \tag{10}$$

则称 C 为凸集。

定义 1.5 (凸函数)**.** 设集合 $C \subset \mathbb{R}^n$ 为非空凸集, 函数 $f : C \to \mathbb{R}$。若对 $\forall \mathbf{x}, \mathbf{y} \in C$，有

$$f(\theta\mathbf{x} + (1-\theta)\mathbf{y}) \leq \theta f(\mathbf{x}) + (1-\theta) f(\mathbf{y}), \quad \theta \in [0, 1] \tag{11}$$

则称 f 为 C 上的凸函数。若上述不等式对 $\mathbf{x} \neq \mathbf{y}$ 严格成立，则称 f 为 C 上的严格凸函数。

定理 1.1 (凸函数的一阶判定条件)**.** *设集合 $C \subset \mathbb{R}^n$ 为非空开凸集，函数 $f : C \to \mathbb{R}$ 可微，则*

1. $f(x)$ 是凸函数当且仅当对 $\forall \mathbf{x}, \mathbf{y} \in C$, 有

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})'(\mathbf{y} - \mathbf{x}) \tag{12}$$

2. $f(\mathbf{x})$ 是严格凸函数当且仅当对 $\forall \mathbf{x}, \mathbf{y} \in C, \mathbf{x} \neq \mathbf{y}$, 有

$$f(\mathbf{y}) > f(\mathbf{x}) + \nabla f(\mathbf{x})'(\mathbf{y} - \mathbf{x}) \tag{13}$$

定理 1.2 (凸函数的二阶判定条件)**.** 设集合 $C \subset \mathbb{R}^n$ 为非空开凸集，函数 $f : C \to \mathbb{R}$ 二阶连续可微，则

1. $f(x)$ 是凸函数当且仅当对 $\forall \mathbf{x} \in C$，*Hessien* 矩阵 $\boldsymbol{H}(\mathbf{x})$ 半正定；
2. 若对 $\forall \mathbf{x} \in C$，*Hessien* 矩阵 $\boldsymbol{H}(\mathbf{x})$ 正定，则 f 是严格凸函数。

1.4 常用矩阵求导公式

关于矩阵及向量求导，这里不加证明地给出以下常用矩阵求导公式，同学们可下来自己利用数学分析与高等代数的知识自己推一遍。

$$\frac{\partial \boldsymbol{\beta}'\mathbf{x}}{\partial \mathbf{x}} = \boldsymbol{\beta} \tag{14}$$

$$\frac{\partial \mathbf{x}'\mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{x} \tag{15}$$

$$\frac{\partial \mathbf{x}'\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = (\mathbf{A} + \mathbf{A}')\,\mathbf{x} \tag{16}$$

评论. 更多关于矩阵运算的知识，可参考“The Matrix Cookbook”(Petersen & Pedersen, 2012)。

2 课堂习题

考虑对正定二次函数 $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}'\mathbf{G}\mathbf{x} + \mathbf{b}'\mathbf{x}$，在点 $\mathbf{x}_k$，求出沿下降方向 $\mathbf{d}_k$ 作精确线搜索的步长 α_k。

证明. 若要给出沿下降方向 $\mathbf{d}_k$ 作精确线搜索的步长 α_k，则应对于 $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{d}_k$，满足

$$\mathbf{d}_k' \nabla f(\mathbf{x}_{k+1}) = 0 \tag{17}$$

由于 $f(\mathbf{x})$ 为正定二次函数，则 $\mathbf{G}$ 为对称正定矩阵，则有 $\mathbf{G}' = \mathbf{G}$，因此

$$\nabla f(\mathbf{x}) = \mathbf{G}\mathbf{x} + \mathbf{b} \tag{18}$$

即

$$\mathbf{d}_k' \nabla f(\mathbf{x}_{k+1}) = \mathbf{d}_k' \left[\mathbf{G}\left(\mathbf{x}_k + \alpha_k \mathbf{d}_k\right) + \mathbf{b}\right] = 0 \tag{19}$$

由于 $\mathbf{G}$ 为正定矩阵，有 $\mathbf{d}_k'\mathbf{G}\mathbf{d}_k > 0$，因此，上式可化简为

$$\alpha_k = -\frac{\mathbf{d}_k'\left(\mathbf{G}\mathbf{x}_k + \mathbf{b}\right)}{\mathbf{d}_k'\mathbf{G}\mathbf{d}_k} = -\frac{\mathbf{d}_k' \nabla f(\mathbf{x}_k)}{\mathbf{d}_k'\mathbf{G}\mathbf{d}_k} \tag{20}$$

□