Lipschitz性与（强）凸性

本文结构

1. 预备知识
2. 定义：
    1. 凸集
    2. 凸函数
    3. 强凸函数
    4. Lipschitz函数
3. Lipschitz函数的性质（在凸集 C 上）
4. 凸函数的性质（在凸集 C 上）
5. 凸函数的性质（在凸集 C 上）
6. 凸函数的性质（在 $\mathbb{R}^n$ 上）

工具

- 向量函数的Taylor公式
- **Lemma 2.1.4.** $\forall 0 \le \alpha \le 1$,

$$\alpha\|x\|^2 + (1-\alpha)\|y\|^2 \ge \alpha(1-\alpha)(\|x\| + \|y\|)^2 \ge \alpha(1-\alpha)\|x \pm y\|^2 \tag{0.1}$$

定义

【凸集】线性空间中一个集合 C 称作**凸的(convex)**，如果 $\forall x, y \in C,\ 0 \le \alpha \le 1$，有 $\alpha x + (1-\alpha)y \in C$

说明：为了方便（利用Taylor公式），接下来的讨论都限制在 $\mathbb{R}^n$ 的子集上，并且假设 f 满足适当的可微性。

我所选择的范数为用 $\mathbb{R}^n$ 的标准内积诱导的向量范数（2-范数）和对应的矩阵范数（谱范数）。这是因为我在证明中用到了向量的内积，所以向量的范数自然就被内积确定了，进而矩阵的范数也就必须是向量范数相应的算子范数。这里没有变化空间。至于用Gram矩阵定义其他内积的情况，因为通过坐标变换就可以还原为标准内积，故也可以不加考虑。

【凸函数】定义在凸集 C 上的实值函数 f 称作**凸的(convex)**，如果 $\forall x, y \in C,\ 0 \le \alpha \le 1$，满足

$$f(\alpha x + (1-\alpha)y) \le \alpha f(x) + (1-\alpha)f(y) \tag{1.1}$$

$$f(y) \ge f(x) + \langle \nabla f(x),\ y - x\rangle \tag{1.2}$$

$$\langle \nabla f(y) - \nabla f(x),\ y - x\rangle \ge 0 \tag{1.3}$$

$$\nabla^2 f(x) \ge 0 \tag{1.4}$$

(1.1)-(1.4)其中之一。

这四种定义得到的凸函数均等价，满足全部4个性质。

> 互相证明：见笔记<函数性质证明>

【强凸函数】定义在凸集 C 上的实值函数 f 称作 **μ-强凸的($\mu-$strongly convex)**，如果

Def 0. 对于某个 $x_0 \in C$，满足

$$f(x) = h(x) + \frac{\mu}{2}\|x - x_0\|^2 \tag{2.0}$$

其中 $h(x)$ 为 C 上一个凸函数。

Def 1-4. $\forall x, y \in C,\ 0 \le \alpha \le 1$，满足

$$\alpha f(x) + (1-\alpha) f(y) - f(\alpha x + (1-\alpha) y) \ge \frac{\alpha(1-\alpha)\mu}{2}\|x - y\|^2 \tag{2.1}$$

$$f(y) \ge f(x) + \langle \nabla f(x),\ y - x\rangle + \frac{\mu}{2}\|y - x\|^2 \tag{2.2}$$

$$\langle \nabla f(y) - \nabla f(x),\ y - x\rangle \ge \mu\|x - y\|^2 \tag{2.3}$$

$$\nabla^2 f(x) \ge \mu I \tag{2.4}$$

(2.1)-(2.4)其中之一。

这五种定义得到的强凸函数均等价，满足全部5个性质。显然，这5个定义都是由凸函数的定义对应而来的。定义0最为清晰，它说明强凸函数其实就是在凸函数的基础上再增加一个二次函数。所以 0-强凸函数就是普通的凸函数。 C 上的 μ-强凸函数族我记为 $\mathcal{S}_\mu(C)$。

> 互相证明：见笔记<函数性质证明>

Remark. 有时候可以放宽对 f 可微的范围的要求，比如(2.4)式只需要在 C 的内部 $\mathring{C}$ 上成立即可。但这并不是我们的重点，所以本文不细致讨论这一问题。

【Lipschitz函数】设 $Q \subseteq \mathbb{R}^n$，令 $\mathcal{C}_L^p(Q)$ 表示满足**Lipschitz性质**

$$\|\nabla^p f(x) - \nabla^p f(y)\| \le L\,\|x - y\|, \quad \forall x, y \in Q \tag{3}$$

的函数族。其中 $p = 0, 1, 2$ 的情况比较常见。

我们也经常考虑同时具有Lipschitz性和凸性的函数，把在凸集 C 上同时满足(2.2)和(3)的函数记为 $\mathcal{S}_{\mu,L}^p(C)$。

Lipschitz函数的性质（在凸集 C 上）

lemma 1.2.2. （用$\nabla^2 f(x)$表示的$\mathcal{C}_L^1(C)$的等价条件）

$$f \in \mathcal{C}_L^1(C) \iff \left\|\nabla^2 f(x)\right\| \le L \iff -LI_n \le \nabla^2 f(x) \le LI_n, \quad \forall x \in C \tag{4}$$

lemma 1.2.3. 若 $f \in \mathcal{C}_L^1(C)$，则 $\forall x, y \in C$ 有

$$|\ f(y) - f(x) - \langle \nabla f(x),\ y - x\rangle\ | \le \frac{L}{2}\|y - x\|^2 \tag{5}$$

lemma 1.2.4. 若 $f \in \mathcal{C}_M^2(C)$，则 $\forall x, y \in C$ 有

$$\left\|\nabla f(y) - \nabla f(x) - \left\langle \nabla^2 f(x),\ y - x\right\rangle\right\| \le \frac{M}{2}\|y - x\|^2 \tag{6.1}$$

$$\left|\ f(y) - f(x) - \langle \nabla f(x),\ y - x\rangle - \frac{1}{2}\left\langle \nabla^2 f(x)(y - x),\ y - x\right\rangle\ \right| \le \frac{M}{6}\|y - x\|^3 \tag{6.2}$$

Cor 1.2.2. 若 $f \in \mathcal{C}_M^2(C)$，则 $\forall x, y \in C$ 有

$$
\begin{aligned}
\nabla^2 f(x) - M\|y-x\| I_n \le \nabla^2 f(y) \le \nabla^2 f(x) + M\|y-x\| I_n \\
-M\|y-x\| I_n \le \nabla^2 f(y) - \nabla^2 f(x) \le M\|y-x\| I_n
\end{aligned}
\tag{6.3}
$$

> 证明思路是用对应阶数的积分形式Taylor公式将 $f(y)$ 在 x 点展开，再利用Lipschitz性放缩。
>
> 详细证明见：笔记<函数性质证明>

凸函数的性质（在凸集 C 上）

Convex + Lipschitz: $\mathcal{S}_{0,L}^1(C)$ 函数的性质

Thm 2.1.6′ （用$\nabla^2 f(x)$表示的$\mathcal{S}_{0,L}^1(C)$的等价条件）

$$
f \in \mathcal{S}_{0,L}^1(\mathbb{R}^n) \iff 0 \le \nabla^2 f(x) \le L I_n, \quad \forall x \in C \tag{7′}
$$

> 证明：$f \in \mathcal{S}_{0,L}^1(\mathbb{R}^n) \iff f \in \mathcal{S}_0(\mathbb{R}^n) \wedge f \in \mathcal{C}_L^1(\mathbb{R}^n)$
>
> 而由(2.4)，$f \in \mathcal{S}_0(\mathbb{R}^n) \iff \nabla^2 f(x) \ge 0$
>
> 由(4)，$f \in \mathcal{C}_L^1(\mathbb{R}^n) \iff -L I_n \le \nabla^2 f(x) \le L I_n$
>
> 故(7)成立 □.

Thm 2.1.5′ 若 $f \in \mathcal{S}_{0,L}^1(C)$，则 $\forall x, y \in C$ 有

$$
0 \le f(y) - f(x) - \langle \nabla f(x),\ y - x \rangle \le \frac{L}{2}\|y-x\|^2 \tag{8.1′}
$$

$$
0 \le \langle \nabla f(y) - \nabla f(x),\ y - x \rangle \le L\|y-x\|^2 \tag{8.2′}
$$

$$
0 \le \alpha f(x) + (1-\alpha) f(y) - f(\alpha x + (1-\alpha) y) \le \frac{\alpha(1-\alpha)L}{2}\|y-x\|^2 \tag{8.3′}
$$

> 证明：$(8.1')$-$(8.3')$左侧的 ≤ 由凸性显然；$(8.1')$右侧的 ≤ 即(5)；
>
> 将$(8.1')$中 x, y 交换可得 $0 \le f(x) - f(y) - \langle \nabla f(y),\ x - y \rangle \le \frac{L}{2}\|y-x\|^2$，左式与$(8.1')$相加即得$(8.2')$；
>
> 要证$(8.3')$右侧，记 $x_\alpha = \alpha x + (1-\alpha) y$，则
>
> $$
> \begin{aligned}
> f(x) - f(x_\alpha) &= \left\langle \int_0^1 \nabla f(x_\alpha + \theta(x - x_\alpha))\, \mathrm{d}\theta, x - x_\alpha \right\rangle \\
> &= \left\langle \int_0^1 \nabla f(x_\alpha + \theta(x - x_\alpha))\, \mathrm{d}\theta,\ (1-\alpha)(x - y) \right\rangle \\
> f(y) - f(x_\alpha) &= \left\langle \int_0^1 \nabla f(x_\alpha + \theta(y - x_\alpha))\, \mathrm{d}\theta, y - x_\alpha \right\rangle \\
> &= \left\langle \int_0^1 \nabla f(x_\alpha + \theta(y - x_\alpha))\, \mathrm{d}\theta,\ \alpha(y - x) \right\rangle
> \end{aligned}
> $$
>
> 所以

> $$
> \begin{aligned}
> &\alpha f(x)+(1-\alpha)f(y)-f(\alpha x+(1-\alpha)y)\\
> &=\alpha\left(f(x)-f(x_\alpha)\right)+(1-\alpha)\left(f(y)-f(x_\alpha)\right)\\
> &=\left\langle\int_0^1\nabla f\left(x_\alpha+\theta(x-x_\alpha)\right)\mathrm{d}\theta,\ \alpha(1-\alpha)(x-y)\right\rangle+\left\langle\int_0^1\nabla f\left(x_\alpha+\theta(y-x_\alpha)\right)\mathrm{d}\theta,\ (1-\alpha)\alpha(y-x)\right\rangle\\
> &=\alpha(1-\alpha)\left\langle y-x,\ \int_0^1\left[\nabla f\left(x_\alpha+\theta(y-x_\alpha)\right)-\nabla f\left(x_\alpha+\theta(y-x_\alpha)\right)\right]\mathrm{d}\theta\right\rangle\\
> &\le\alpha(1-\alpha)\|y-x\|\int_0^1 L\|y-x\|\theta\mathrm{d}\theta=\frac{\alpha(1-\alpha)L}{2}\|y-x\|^2\quad\square.
> \end{aligned}
> $$

Strongly Convex: $\mathcal{S}_\mu(C)$ 函数的性质

Thm 2.1.10′ 若 $f\in\mathcal{S}_\mu(C)$，则 $\forall x,y\in C$ 有

$$
\mu\|x-y\|\le\|\nabla f(x)-\nabla f(y)\| \tag{9.1′}
$$

> 证明：由(2.3)
>
> $$
> \|\nabla f(y)-\nabla f(x)\|\,\|y-x\|\ge\langle\nabla f(y)-\nabla f(x),\ y-x\rangle\ge\mu\|x-y\|^2\quad\square.
> $$