最优化导论第三次作业题

1. 设 $f(x)$ 为凸函数。证明： $f(x)$ 为凸函数的充要条件是对任意的 $x, y \in \mathbb{R}^n$，一元函数

$$\varphi(\alpha) = f(x + \alpha y)$$

是关于 α 的凸函数。

证明:（必要性）

设 $\lambda_1 \geq 0, \lambda_2 \geq 0$，且 $\lambda_1 + \lambda_2 = 1$。
由 $\varphi(\alpha)$ 的定义和 $f(x)$ 的凸性，有：

$$\begin{aligned}
\varphi(\lambda_1\alpha_1 + \lambda_2\alpha_2) &= f(x + (\lambda_1\alpha_1 + \lambda_2\alpha_2)y) \\
&= f(\lambda_1(x + \alpha_1 y) + \lambda_2(x + \alpha_2 y)) \\
&\leq \lambda_1 f(x + \alpha_1 y) + \lambda_2 f(x + \alpha_2 y) \\
&= \lambda_1\varphi(\alpha_1) + \lambda_2\varphi(\alpha_2)
\end{aligned}$$

由定义知 $\varphi(\alpha)$ 是凸函数。

（充分性）

任取 $x, y \in \mathbb{R}^n$，记

$$\varphi(\alpha) = f(x + \alpha(y - x))$$

则

$$\begin{aligned}
f(\lambda_1 x + \lambda_2 y) &= f(x + \lambda_2(y - x)) \\
&= \varphi(\lambda_2) \\
&\leq \lambda_1\varphi(0) + \lambda_2\varphi(1) \\
&= \lambda_1 f(x) + \lambda_2 f(y)
\end{aligned}$$

故知 $f(x)$ 为凸函数。

2. 设 $f : \mathbb{R}^n \to \mathbb{R}$ 为连续函数。 证明：f 是凸函数当且仅当对于任意的 $x, y \in \mathbb{R}^n$， 其在连线上取值的平均值不超过两端点函数值的平均值，即：

$$\int_0^1 f(x + \lambda(y - x))\, d\lambda \leq \frac{f(x) + f(y)}{2}.$$

证明

(充分性)
若 f 为凸函数，则根据 Jensen 不等式，对于任意 $0 \leq \lambda \leq 1$，有：

$$f(x + \lambda(y - x)) \leq f(x) + \lambda(f(y) - f(x)).$$

对两边在 $[0, 1]$ 上积分：

$$\int_0^1 f(x + \lambda(y - x))\, d\lambda \leq \int_0^1 (f(x) + \lambda(f(y) - f(x)))\, d\lambda = \frac{f(x) + f(y)}{2}.$$

因此，若 f 为凸函数，则该不等式成立。

(必要性)
现在证明反向，即若上述积分不等式成立，则 f 为凸函数。

假设 f 不是凸函数，则存在 $x, y \in \mathbb{R}^n$ 及某个 $\theta_0 \in (0,1)$，使得：

$$f(\theta_0 x + (1-\theta_0)y) > \theta_0 f(x) + (1-\theta_0) f(y).$$

定义函数：

$$F(\theta) = f(\theta x + (1-\theta)y) - [\theta f(x) + (1-\theta) f(y)],$$

由于 f 连续，F 也是连续函数。
显然 $F(0) = F(1) = 0$，且在 θ_0 处有 $F(\theta_0) > 0$。

令 α 为 F 在 θ_0 左侧的最大零点，β 为 F 在 θ_0 右侧的最小零点。

定义 $u = \alpha x + (1-\alpha)y, \quad v = \beta x + (1-\beta)y.$

在区间 (α, β) 上，$F(\theta) > 0$，即：

$$f(\theta x + (1-\theta)y) > \theta f(x) + (1-\theta) f(y).$$

由此可得：

$$f(\theta u + (1-\theta)v) > \theta f(u) + (1-\theta) f(v), \quad \forall \theta \in (0,1).$$

对 θ 从 0 到 1 积分：

$$\int_0^1 f(u + \theta(v-u))\, d\theta > \int_0^1 \big(f(u) + \theta(f(v) - f(u))\big)\, d\theta = \frac{f(u) + f(v)}{2}.$$

这意味着：在区间 $[u, v]$ 上，f 的平均值大于两端点处的平均值，
与题设假设矛盾。

因此，命题得证。

3. 设 $f : \mathbb{R} \to \mathbb{R}$ 为凸函数，且 $\mathbb{R}_+ \subseteq \operatorname{dom} f$。定义其"滑动平均"函数

$$F(x) = \frac{1}{x} \int_0^x f(t)\, dt, \qquad \operatorname{dom} F = \mathbb{R}_{++}.$$

证明 F 为凸函数。（可假设 f 可微。）

证明:

由积分求导法则，

$$F'(x) = -\frac{1}{x^2} \int_0^x f(t)\, dt + \frac{1}{x} f(x).$$

再求一次导数得

$$
\begin{aligned}
F''(x) &= \frac{2}{x^3}\int_0^x f(t)\,dt - \frac{2}{x^2}f(x) + \frac{1}{x}f'(x) \\
&= \frac{2}{x^3}\int_0^x \big(f(t)-f(x)\big)\,dt + \frac{1}{x}f'(x) \\
&= \frac{2}{x^3}\int_0^x \big(f(t)-f(x)-f'(x)(t-x)\big)\,dt.
\end{aligned}
$$

由于 f 凸且可微，对任意 $x, t \in \operatorname{dom} f$ 有不等式

$$f(t) \ \geq \ f(x) + f'(x)(t-x).$$

故被积函数

$$f(t) - f(x) - f'(x)(t-x) \ \geq \ 0,$$

从而

$$F''(x) \ = \ \frac{2}{x^3}\int_0^x \underbrace{\big(f(t)-f(x)-f'(x)(t-x)\big)}_{\geq 0}\,dt \ \geq \ 0.$$

因此 F 在 $\mathbb{R}_{++}$ 上二阶导非负，故 F 为凸函数。证毕。

4. 判断下列函数在给定定义域上是否为凸函数：

(a) $f(x) = e^x - 1$，定义域 $\mathbb{R}$.

(b) $f(x_1, x_2) = x_1 x_2$，定义域 $\mathbb{R}^2_{++}$.

(c) $f(x_1, x_2) = 1/(x_1 x_2)$，定义域 $\mathbb{R}^2_{++}$.

(d) $f(x_1, x_2) = x_1/x_2$，定义域 $\mathbb{R}^2_{++}$.

(e) $f(x_1, x_2) = x_1^2/x_2$，定义域 $\mathbb{R} \times \mathbb{R}_{++}$.

(f) $f(x_1, x_2) = x_1^{\alpha} x_2^{1-\alpha}$，其中 $0 \leq \alpha \leq 1$，定义域 $\mathbb{R}^2_{++}$.

解:

(a) 凸。因为 $f''(x) = e^x > 0$（处处严格正），故严格凸。

(b) 非凸。Hessian 为

$\nabla^2 f = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$，既非正半定也非负半定，故非凸。

(c) 凸。
$\nabla^2 f = \dfrac{1}{x_1 x_2}\begin{bmatrix} 2/x_1^2 & 1/(x_1 x_2) \\ 1/(x_1 x_2) & 2/x_2^2 \end{bmatrix} \succeq 0$，故凸。

(d) 非凸。

$\nabla^2 f = \begin{bmatrix} 0 & -1/x_2^2 \\ -1/x_2^2 & 2x_1/x_2^3 \end{bmatrix}$ 不为正半定（亦非负半定），故非凸。

(e) 凸。

$\nabla^2 f = \dfrac{2}{x_2}\begin{bmatrix} 1 & -x_1/x_2 \\ -x_1/x_2 & x_1^2/x_2^2 \end{bmatrix} \succeq 0$，故凸。

(f) 非凸（为凹）。

Hessian $\nabla^2 f \preceq 0$（当 $0 \le \alpha \le 1$），因此函数是凹函数而非凸函数。

5. 证明以下函数在其定义域上是凸函数，可以使用复合规则：

(a)

$$f(x) = -\log\Big(-\log\Big(\sum_{i=1}^{m} e^{a_i^\top x + b_i}\Big)\Big), \quad \mathrm{dom} f = \Big\{x \mid \sum_{i=1}^{m} e^{a_i^\top x + b_i} < 1\Big\}.$$

证明：
令 $g(x) = \log \sum_i e^{a_i^\top x + b_i}$。由 `log-sum-exp` 的性质，$g$ 为凸函数。 取 $h(y) = -\log y$，其在 $\mathbb{R}_{++}$ 上凸且单调递减。 因此

$$f(x) = h(-g(x))$$

是"凸且递减函数"与"凹函数"的复合，仍为凸函数。

(b)

$$f(x,u,v) = -\sqrt{uv - x^\top x}, \quad \mathrm{dom} f = \{(x,u,v) \mid u > 0, v > 0, uv > x^\top x\}.$$

证明： 我们可将 f 写成

$$f(x,u,v) = -\sqrt{u\big(v - x^T x/u\big)}.$$

令

$$h(x_1, x_2) = -\sqrt{x_1 x_2},$$

则 h 在 $\mathbb{R}_{++}^2$ 上是凸函数，且对每个自变量单调不增。
再令

$$g_1(u,v,x) = u, \qquad g_2(u,v,x) = v - \frac{x^T x}{u}.$$

其中 g_1 为仿射函数（既凸又凹），而由于 $x^T x/u$ 是凸的，故
g_2 为凹函数。于是

$$f(u,v,x) = h\big(g_1(u,v,x),\, g_2(u,v,x)\big)$$

为凸函数（凸且分别非增的外层函数与凹的内层函数的复合仍为凸）。

(c)

$$f(x,u,v) = -\log(uv - x^\top x), \quad \mathrm{dom} f = \{(x,u,v) \mid u > 0, v > 0, uv > x^\top x\}.$$

证明:
分解:

$$f(x,u,v) = -\log u - \log\Big(v - \frac{x^\top x}{u}\Big).$$

$-\log u$ 为凸函数; $v - \frac{x^\top x}{u}$ 为凹函数（因为 $\frac{x^\top x}{u}$ 凸且取负变凹）; $-\log(\cdot)$ 是凸且递减函数;
因此第二项也是凸函数。 两项相加 ⇒ f 凸。

(d)

$$f(x,t) = -\big(t^p - \|x\|_p^p\big)^{1/p}, \quad p>1, \quad \mathrm{dom} f = \{(x,t) \mid t \ge \|x\|_p\}.$$

证明:
改写为:

$$f(x,t) = -\left(t - \frac{\|x\|_p^p}{t^{p-1}}\right)^{1/p}.$$

设

$$g_1(t) = t^{1-1/p}, \quad g_2(x,t) = t - \frac{\|x\|_p^p}{t^{p-1}},$$

两者均为凹函数。
定义 $h(y_1,y_2) = -y_1 y_2$，其对每个自变量都是凸且递减。 由复合规则，$f = h(g_1,g_2)$ 是凸函数。

(e)

$$f(x,t) = -\log\big(t^p - \|x\|_p^p\big), \quad p>1, \quad \mathrm{dom} f = \{(x,t) \mid t > \|x\|_p\}.$$

证明:
展开:

$$f(x,t) = -\log t^{p-1} - \log\Big(t - \frac{\|x\|_p^p}{t^{p-1}}\Big) = -(p-1)\log t - \log\Big(t - \frac{\|x\|_p^p}{t^{p-1}}\Big).$$

第一项凸。 第二项是“递减凸函数 $-\log(\cdot)$” 与“凹函数 $t - \|x\|_p^p/t^{p-1}$”的复合，仍为凸函数。 两项相加 ⇒ f 凸。

注：常用结论:

- `log-sum-exp` 函数 $\log\sum_i e^{z_i}$ 是凸的;
- $-\log(y)$ 在 $\mathbb{R}_{++}$ 上凸且严格递减;
- $\sqrt{xy}$ 在 $\mathbb{R}_{++}^2$ 上凹;
- $\frac{x^\top x}{u}$ 在 $\{u>0\}$ 上对 (x,u) 凸;
- $\|x\|_p^p/u^{p-1}$ 在 $\{u>0\}$ 上对 (x,u) 凸;
- 若 h 凸且非增，而 g 凹，则 $h \circ g$ 凸。

6. 设 $f:\mathbb{R}^n\to\mathbb{R}$ 是定义在整个 $\mathbb{R}^n$ 上的凸函数。
若存在一个有限划分

$$\mathbb{R}^n = X_1\cup X_2\cup\cdots\cup X_L,$$

其中每个 X_i 的内部非空，且 $\operatorname{int} X_i\cap \operatorname{int} X_j=\emptyset$ （当 $i\neq j$ 时），
并且在每个子集 X_i 上，f 都是仿射函数：$f(x)=a_i^\top x+b_i,\quad x\in X_i.$

证明： $f(x)=\max_{i=1,\dots,L}(a_i^\top x+b_i).$

证明:

由 Jensen 不等式，对任意 $x,y\in\operatorname{dom} f$ 及 $t\in[0,1]$，有:

$$f(y+t(x-y))\le f(y)+t(f(x)-f(y)).$$

移项可得:

$$f(x)\ge f(y)+\frac{f(y+t(x-y))-f(y)}{t}.$$

设 $x\in X_i$，取任意 $y\in\operatorname{int} X_j$， 并取足够小的 $t>0$，使得 $y+t(x-y)\in X_j$。

由于在 X_i,X_j 上 f 均为仿射函数:

$$f(x)=a_i^\top x+b_i,\quad f(y+t(x-y))=a_j^\top(y+t(x-y))+b_j.$$

代入不等式得:

$$a_i^\top x+b_i\ge a_j^\top y+b_j+\frac{a_j^\top(y+t(x-y))+b_j-a_j^\top y-b_j}{t}=a_j^\top x+b_j.$$

上式对任意 $j=1,\dots,L$ 都成立，因此：$a_i^\top x+b_i\ge \max_{j=1,\dots,L}(a_j^\top x+b_j).$

故而:

$$f(x)=a_i^\top x+b_i=\max_{j=1,\dots,L}(a_j^\top x+b_j).$$

7. 设 $f:\mathbb{R}^n\to\mathbb{R}$ 为凸函数，定义其透视函数（perspective function）为

$$g(x,t)=t\,f(x/t),\quad \text{定义域 } \operatorname{dom} g=\{(x,t)\mid x/t\in\operatorname{dom} f,\ t>0\}.$$

证明:

(a) $\operatorname{dom} g$ 是凸集;

(b) 对任意 $(x,t),(y,s)\in\operatorname{dom} g$，以及 $0\le\theta\le1$，成立:

$$g(\theta x+(1-\theta)y,\ \theta t+(1-\theta)s)\le \theta g(x,t)+(1-\theta)g(y,s).$$

证明: (a) 取任意 $(x,t),(y,s)\in \operatorname{dom} g$ 与 $\theta\in[0,1]$。则 $t>0, s>0$ 且 $x/t,\ y/s\in \operatorname{dom} f$。记

$$\lambda := \frac{\theta t}{\theta t+(1-\theta)s}\in[0,1],$$

则有

$$\frac{\theta x+(1-\theta)y}{\theta t+(1-\theta)s}=\lambda\frac{x}{t}+(1-\lambda)\frac{y}{s}.$$

由于 $\operatorname{dom} f$ 为凸集，$\lambda(x/t)+(1-\lambda)(y/s)\in \operatorname{dom} f$。
又 $\theta t+(1-\theta)s>0$，故

$$\big(\theta x+(1-\theta)y,\ \theta t+(1-\theta)s\big)\in \operatorname{dom} g.$$

于是 $\operatorname{dom} g$ 对任意凸组合封闭，因而为凸集。

(b) 假设 $s,t>0$，且 $x/t, y/s\in \operatorname{dom} f$，令 $0\le\theta\le 1$。
需证:

$$g(\theta x+(1-\theta)y,\ \theta t+(1-\theta)s)\le \theta g(x,t)+(1-\theta)g(y,s).$$

由定义:

$$g(\theta x+(1-\theta)y,\ \theta t+(1-\theta)s)=(\theta t+(1-\theta)s)\,f\!\left(\frac{\theta x+(1-\theta)y}{\theta t+(1-\theta)s}\right).$$

注意到:

$$\frac{\theta x+(1-\theta)y}{\theta t+(1-\theta)s}=\frac{\theta t}{\theta t+(1-\theta)s}\cdot\frac{x}{t}+\frac{(1-\theta)s}{\theta t+(1-\theta)s}\cdot\frac{y}{s}.$$

令权重

$$\lambda=\frac{\theta t}{\theta t+(1-\theta)s},\quad 1-\lambda=\frac{(1-\theta)s}{\theta t+(1-\theta)s}.$$

显然 $0\le\lambda\le 1$，且 $\lambda+(1-\lambda)=1$。

由于 f 为凸函数，利用 Jensen 不等式:

$$f\!\left(\frac{\theta x+(1-\theta)y}{\theta t+(1-\theta)s}\right)\le\frac{\theta t}{\theta t+(1-\theta)s}f(x/t)+\frac{(1-\theta)s}{\theta t+(1-\theta)s}f(y/s).$$

两边同乘以 $(\theta t+(1-\theta)s)$，得:

$$\begin{aligned}g(\theta x+(1-\theta)y,\ \theta t+(1-\theta)s)&\le \theta t f(x/t)+(1-\theta)s f(y/s)\\&=\theta g(x,t)+(1-\theta)g(y,s).\end{aligned}$$