最优化导论　第三次作业

课本习题：　《Convex Optimization》3.5, 3.6, 3.7, 3.8, 3.10, 3.11, 3.12, 3.16

3.5 [RV73, 第 22 页] 凸函数的滑动平均。假设 $f : \mathrm{R} \to \mathrm{R}$ 是凸函数，且 $\mathrm{R}_+ \subseteq \operatorname{dom} f$。证明其滑动平均 F 是凸的，其中 F 定义为

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

并假设 f 是可微的。

证明　F 是可微的，其导数为

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

由以下事实可得 F 的凸性：

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

对所有 $x, t \in \operatorname{dom} f$ 都成立，这意味着 $F''(x) \geq 0$。

3.6 函数与上图。一个函数的上图何时是半空间？一个函数的上图何时是凸锥？一个函数的上图何时是多面体？

解　如果函数分别是仿射的、正齐次的（即满足 $f(\alpha x) = \alpha f(x)$ 对于 $\alpha \geq 0$ 成立）、以及分段仿射的。

3.7 假设 $f : \mathrm{R}^n \to \mathrm{R}$ 是一个凸函数，且 $\operatorname{dom} f = \mathrm{R}^n$，在 R^n 上有上界。证明 f 是常数函数。

解 假设 f 不是常数，即存在 x, y 使得 $f(x) < f(y)$。定义函数

$$g(t) = f(x + t(y - x))$$

则 $g(t)$ 是凸函数，且 $g(0) < g(1)$。根据 Jensen 不等式，有

$$g(1) \leq \frac{t-1}{t} g(0) + \frac{1}{t} g(t)$$

对于所有 $t > 1$ 成立，因此

$$g(t) \geq t g(1) - (t-1) g(0) = g(0) + t(g(1) - g(0))$$

所以当 $t \to \infty$ 时，g 将无界增长。这与我们假设 f 有上界相矛盾。

3.8 凸性的二阶条件。证明一个二次可微函数 f 是凸的，当且仅当其定义域是凸集，且对于所有 $x \in \operatorname{dom} f$，都有 $\nabla^2 f(x) \succeq 0$。提示。首先考虑 $f : \mathrm{R} \to \mathrm{R}$ 的情况。可以使用凸性的一级条件（见第 70 页证明）。

解 首先假设 $n = 1$。假设 $f : \mathrm{R} \to \mathrm{R}$ 是凸函数。令 $x, y \in \operatorname{dom} f$ 且 $y > x$。根据一级条件，

$$f'(x)(y - x) \leq f(y) - f(x) \leq f'(y)(y - x)$$

将右侧减去左侧并除以 $(y - x)^2$，得到

$$\frac{f'(y) - f'(x)}{y - x} \geq 0$$

令 $y \to x$，得到 $f''(x) \geq 0$。

反之，假设对于所有 $z \in \operatorname{dom} f$，有 $f''(z) \geq 0$。考虑任意两点 $x, y \in \operatorname{dom} f$ 且 $x < y$，则有

$$\begin{aligned}
0 &\leq \int_x^y f''(z)(y - z)\, dz \\
&= \left(f'(z)(y - z)\right)\Big|_{z=x}^{z=y} + \int_x^y f'(z)\, dz \\
&= -f'(x)(y - x) + f(y) - f(x)
\end{aligned}$$

即 $f(y) \geq f(x) + f'(x)(y - x)$。这表明 f 是凸的。

对于 $n > 1$ 的情况，我们注意到一个函数是凸的，当且仅当它在所有直线上都是凸的，即对于所有 $x_0 \in \operatorname{dom} f$ 和所有 v，函数 $g(t) = f(x_0 + tv)$ 在 t 上是凸的。因此，f 是凸的，当且仅当

$$g''(t) = v^T \nabla^2 f(x_0 + tv)\, v \geq 0$$

对于所有 $x_0 \in \operatorname{dom} f, v \in \mathrm{R}^n$，以及满足 $x_0 + tv \in \operatorname{dom} f$ 的 t。换句话说，对于所有 $x \in \operatorname{dom} f$，$\nabla^2 f(x) \succeq 0$ 是必要且充分条件。

3.10 Jensen 不等式的扩展。Jensen 不等式的一种解释是，随机化或抖动会有害，即会提升凸函数的平均值：对于凸函数 f 和零均值随机变量 v，有 $\mathrm{E} f(x_0 + v) \geq f(x_0)$。这引出了以下猜

想：若 f_0 是凸函数，则 v 的方差越大，$\mathrm{E}f(x_0+v)$ 越大。(a) 给出一个反例证明该猜想是错误的。找出零均值随机变量 v 和 w，使得 $\mathrm{var}(v) > \mathrm{var}(w)$，一个凸函数 f，以及一个点 x_0，满足 $\mathrm{E}f(x_0+v) < \mathrm{E}f(x_0+w)$。(b) 当 v 和 w 是彼此的缩放版本时，该猜想是成立的。证明当 f 是凸函数且 v 是零均值时，$\mathrm{E}f(x_0+tv)$ 在 $t \geq 0$ 时单调递增。

解 (a) 定义 $f: \mathrm{R} \to \mathrm{R}$ 为

$$f(x) = \begin{cases} 0, & x \leq 0 \\ x, & x > 0 \end{cases}$$

设 $x_0 = 0$，并定义标量随机变量

$$w = \begin{cases} 1 & \text{with probability } 1/2 \\ -1 & \text{with probability } 1/2 \end{cases} \qquad v = \begin{cases} 4 & \text{with probability } 1/10 \\ -4/9 & \text{with probability } 9/10 \end{cases}$$

w 和 v 均为零均值，并且

$$\mathrm{var}(v) = 16/9 > 1 = \mathrm{var}(w)$$

然而，

$$\mathrm{E}f(v) = 2/5 < 1/2 = \mathrm{E}f(w)$$

(b) 对于固定的 v，$f(x_0+tv)$ 关于 t 是凸的，因此若 v 是随机变量，则 $g(t) = \mathrm{E}f(x_0+tv)$ 是关于 t 的凸函数。根据 Jensen 不等式，

$$g(t) = \mathrm{E}f(x_0+tv) \geq f(x_0) = g(0)$$

现在考虑两点 a, b，使得 $0 < a < b$。若 $g(b) < g(a)$，则

$$\frac{b-a}{b}g(0) + \frac{a}{b}g(b) < \frac{b-a}{b}g(a) + \frac{a}{b}g(a) = g(a)$$

这与 Jensen 不等式矛盾。因此我们必须有 $g(b) \geq g(a)$。

3.11 单调映射。一个函数 $\psi: \mathrm{R}^n \to \mathrm{R}^n$ 称为单调的，如果对于所有 $x, y \in \mathrm{dom}\,\psi$，满足

$$(\psi(x) - \psi(y))^T(x-y) \geq 0$$

（注意，这里的 “单调” 与 §3.6.1 中给出的定义不同。这两种定义在文献中均被广泛使用。）假设 $f: \mathrm{R}^n \to \mathrm{R}$ 是一个可微的凸函数。证明其梯度 ∇f 是单调的。其逆命题是否成立，即每一个单调映射是否都是某个凸函数的梯度?

解 f 的凸性意味着

$$f(x) \geq f(y) + \nabla f(y)^T(x-y), \quad f(y) \geq f(x) + \nabla f(x)^T(y-x)$$

对于任意 $x, y \in \mathbf{dom}\, f$ 成立。将这两个不等式结合可得

$$(\nabla f(x) - \nabla f(y))^T(x-y) \geq 0$$

这表明 ∇f 是单调的。逆命题通常不成立。作为反例，考虑

$$\psi(x) = \begin{bmatrix} x_1 \\ x_1/2 + x_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1/2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

ψ 是单调的，因为

$$(x-y)^T \begin{bmatrix} 1 & 0 \\ 1/2 & 1 \end{bmatrix} (x-y) = (x-y)^T \begin{bmatrix} 1 & 1/4 \\ 1/4 & 1 \end{bmatrix} (x-y) \geq 0$$

对于所有 x, y 成立。然而，不存在一个函数 $f : \mathrm{R}^2 \to \mathrm{R}$ 使得 $\psi(x) = \nabla f(x)$，因为这样的函数必须满足

$$\frac{\partial^2 f}{\partial x_1 \partial x_2} = \frac{\partial \psi_1}{\partial x_2} = 0, \quad \frac{\partial^2 f}{\partial x_1 \partial x_2} = \frac{\partial \psi_2}{\partial x_1} = 1/2.$$

3.12 假设 $f : \mathrm{R}^n \to \mathrm{R}$ 是凸函数，$g : \mathrm{R}^n \to \mathrm{R}$ 是凹函数，且 $\mathbf{dom}\, f = \mathbf{dom}\, g = \mathrm{R}^n$，对所有 x，有 $g(x) \leq f(x)$。证明存在一个仿射函数 h 使得对所有 x，$g(x) \leq h(x) \leq f(x)$。换句话说，如果一个凹函数 g 是凸函数 f 的下界估计函数，那么我们可以在 f 和 g 之间拟合一个仿射函数。

解 首先注意到 int epi f 是非空的（因为 $\mathbf{dom}\, f = \mathrm{R}^n$），并且不会与 hypo g 相交（因为对于 $(x,t) \in \mathbf{int}\,\mathbf{epi}\, f$ 有 $f(x) < t$，而对于 $(x,t) \in \mathbf{hypo}\, g$ 有 $t \geq g(x)$）。因此，这两个集合可以被一个超平面分开，即存在 $a \in \mathrm{R}^n, b \in \mathrm{R}$（二者不同时为零）以及 $c \in \mathrm{R}$，使得

$$a^T x + bt \geq c \geq a^T y + bv$$

当 $t > f(x)$ 且 $v \leq g(y)$ 时成立。我们必须有 $b \neq 0$，否则该条件将简化为对于所有 x 和 y 满足 $a^T x \geq a^T y$，这仅在 $a = 0$ 时可能成立。选择 $x = y$，并利用 $f(x) \geq g(x)$ 的事实，我们还可以得出 $b > 0$。

现在我们将分离超平面条件应用于点 $(x,t) \in \mathbf{int}\,\mathbf{epi}\, f$，以及 $(y,v) = (x, g(x)) \in \mathbf{hypo}\, g$，并得到

$$a^T x + bt \geq c \geq a^T x + bg(x)$$

两边除以 b 得到

$$t \geq \left(c - a^T x\right)/b \geq g(x)$$

对于所有 $t > f(x)$。因此，仿射函数 $h(x) = \left(c - a^T x\right)/b$ 位于 f 和 g 之间。

3.16 对下列每个函数，确定它是凸的、凹的、拟凸的，还是拟凹的。(a) $f(x)=e^{x}-1$ 在 R 上。

解答。该函数是严格凸的，因此是拟凸的。它也是拟凹的，但不是凹的。

(b) $f(x_1, x_2)=x_1 x_2$ 在 R_{++}^{2} 上。

解答。f 的 Hessian 矩阵为

$$\nabla^{2} f(x)=\left[\begin{array}{ll}0 & 1 \\ 1 & 0\end{array}\right]$$

该矩阵既不是半正定的也不是半负定的。因此，f 既不是凸的也不是凹的。它是拟凹的，因为其超水平集

$$\left\{(x_1, x_2) \in \mathrm{R}_{++}^{2} \mid x_1 x_2 \geq \alpha\right\}$$

是凸的。它不是拟凸的。

(c) $f(x_1, x_2)=\frac{1}{x_1 x_2}$ 在 R_{++}^{2} 上。

解 f 的 Hessian 矩阵为

$$\nabla^{2} f(x)=\frac{1}{x_1 x_2}\left[\begin{array}{cc}\frac{2}{x_1^{2}} & \frac{1}{x_1 x_2} \\ \frac{1}{x_1 x_2} & \frac{2}{x_2^{2}}\end{array}\right] \succeq 0$$

因此，f 是凸的且是拟凸的。它既不是拟凹的，也不是凹的。

(d) $f(x_1, x_2)=\frac{x_1}{x_2}$ 在 R_{++}^{2} 上。

解答。f 的 Hessian 矩阵为

$$\nabla^{2} f(x)=\left[\begin{array}{cc}0 & -\frac{1}{x_2^{2}} \\ -\frac{1}{x_2^{2}} & \frac{2 x_1}{x_2^{3}}\end{array}\right]$$

该矩阵既不是半正定的也不是半负定的。因此，f 既不是凸的也不是凹的。它是拟凸的和拟凹的（即拟线性的），因为其子水平集和超水平集是半空间。

(e) $f(x_1, x_2)=\frac{x_1^{2}}{x_2}$ 在 $\mathrm{R} \times \mathrm{R}_{++}$ 上。

解答。f 是凸的，如第 72 页所述（参见图 3.3）。这可以通过计算 Hessian 来轻松验证：

$$\nabla^{2} f(x)=\left[\begin{array}{cc}\frac{2}{x_2} & -\frac{2 x_1}{x_2^{2}} \\ -\frac{2 x_1}{x_2^{2}} & \frac{2 x_1^{2}}{x_2^{3}}\end{array}\right]=\left(\frac{2}{x_2}\right)\left[\begin{array}{c}1 \\ -\frac{2 x_1}{x_2}\end{array}\right]\left[\begin{array}{ll}1 & -\frac{2 x_1}{x_2}\end{array}\right] \succeq 0.$$

因此，f 是凸的且是拟凸的。它不是凹的或拟凹的（见图）。

(f) $f(x_1, x_2)=x_1^{\alpha} x_2^{1-\alpha}$，其中 $0 \leq \alpha \leq 1$，在 R_{++}^{2} 上。

解答。f 是凹的和拟凹的。f 的 Hessian 矩阵为

$$
\begin{aligned}
\nabla^2 f(x) &= \begin{bmatrix} \alpha(\alpha-1)x_1^{\alpha-2}x_2^{1-\alpha} & \alpha(1-\alpha)x_1^{\alpha-1}x_2^{-\alpha} \\ \alpha(1-\alpha)x_1^{\alpha-1}x_2^{-\alpha} & (1-\alpha)(-\alpha)x_1^{\alpha}x_2^{-\alpha-1} \end{bmatrix} \\
&= \alpha(1-\alpha)x_1^{\alpha}x_2^{1-\alpha} \begin{bmatrix} -\frac{1}{x_1^2} & \frac{1}{x_1x_2} \\ \frac{1}{x_1x_2} & -\frac{1}{x_2^2} \end{bmatrix} \\
&= -\alpha(1-\alpha)x_1^{\alpha}x_2^{1-\alpha} \begin{bmatrix} \frac{1}{x_1} \\ -\frac{1}{x_2} \end{bmatrix} \begin{bmatrix} \frac{1}{x_1} \\ -\frac{1}{x_2} \end{bmatrix}^T \\
&\preceq 0.
\end{aligned}
$$

因此，f 既不是凸的，也不是拟凸的。