最优化理论与算法

1 第一次作业

1.1 证明保凸运算

1. 凸集的交集仍为凸集

证明. 原命题等价于证明:n 个（$n \geq 0, n \in \mathbb{N}_+$）凸集的交集仍然是凸集。

1. 当 $n = 1$ 时：$\cap$ 运算是二元运算符，不考虑这种情况。
2. 当 $n = 2$ 时：不妨设 A, B 为两个凸集。
   当 $A \cap B = \emptyset$ 时，由凸集的定义，$\emptyset$ 也是一个凸集。
   当 $A \cap B \neq \emptyset$ 时，设 $x_1, x_2 \in A \cap B$, 则 x_1, x_2 也满足 $x_1 \in A$ 且 $x_1 \in B$, $x_2 \in A$ 且 $x_2 \in B$。因为 A, B 均为凸集，由凸集定义:
   $$\theta x_1 + (1-\theta)x_2 \in A\text{且}\theta x_1 + (1-\theta)x_2 \in B$$
   故：
   $$\theta x_1 + (1-\theta)x_2 \in A \cap B$$
3. 假设 $n = k$ 时原命题成立。也即存在凸集 $A_1, A_2 \ldots A_k$ 使得。
   $$\bigcap_i^k A_i\text{为凸集}$$
   当 $n = k+1$ 时，$A_1, A_2 \ldots A_{k+1}$ 为凸集，则：
   $$\bigcap_i^{k+1} A_i = \bigcap_i^k A_i \cap A_{k+1}$$
   由 $n = 2, n = k$ 时原命题成立可知，$\bigcap_i^k A_i$ 为凸集，A_{k+1} 也为凸集，故 $\bigcap_i^k A_i \cap A_{k+1}$ 也为凸集。故 $n = k+1$ 时原命题也成立。

由数学归纳法知：原命题成立。 □

2. C 是凸集, 则 $aC + b$ 是凸集

证明. 根据题意可知：

$$S = aC + b = \{y|y = ax + b, x \in C\}$$

对 $\forall y_1, y_2 \in S$, $x_1, x_2 \in C$ 及 $\forall t \in [0,1]$, 根据集合 S 的定义可得：

$$y_1 = ax_1 + b, y_2 = ax_2 + b$$

$$ty_1 + (1-t)y_2 = t(ax_1 + b) + (1-t)(ax_2 + b) = a(tx_1 + (1-t)x_2) + b$$

$\because C$ 为凸集, 根据凸集的定义有：$tx_1 + (1-t)x_2 \in C, t \in [0,1]$

$\therefore ty_1 + (1-t)y_2 \in S, t \in [0,1]$, 则 S 为凸集。 □

3. C 是凸集, $f(x) = Ax + b$, 则仿射项 $f(C)$ 是凸集

证明. 对 $\forall x_1, x_2 \in C$, 由于 C 为凸集，则 $\forall t \in [0,1]$

$$tx_1 + (1-t)x_2 \in C$$

对凸集 C 进行仿射变换得到 $AC + b$, 记为集合 S, 由仿射变换的定义可知：

$$x_1, x_2 \in C, Ax_1 + b, Ax_2 + b \in S$$

$$\therefore f(tx_1 + (1-t)x_2) = A(tx_1 + (1-t)x_2) + b = t(Ax_1 + b) + (1-t)(Ax_2 + b) \in S$$

所以 S 为凸集, 该命题得以证明。 □

4. D 是凸集, $f(x) = Ax + b$, 则原项 $f^{-1}(D)$ 是凸集

证明. 由题意知：

$$f(x) = Ax + b$$

将其变换, 对 x 进行求解得：

$$x = A^{-1}(f(x) - b)$$

等价于：

$$x = A^{-1}f(x) - A^{-1}b$$

令 $x = f(y), f(x) = y, A^{-1} = B, -A^{-1}b = c$, 则可将逆变换的式子写作：

$$f(y) = By + c$$

对比 $f(x) = Ax + b$, 可知逆变换可以写作仿射变换的形式, 由（1）所证明的仿射变换的保凸运算, 可得逆变换后仍保凸, 所以该命题得以证明。 □

1.2 证明无穷范数满足三性质: 正定性, 齐次性和次可加性

证明. 1. 正定性：

$\because ||x||_\infty = \max_i |x_i|$, 根据绝对值的性质,$|x| \geq 0$,

$\therefore ||x||_\infty \geq 0$, 若取等, 当且仅当 $x = 0$ 时成立。

即无穷范数具有正定性得以证明。

2. 齐次性：

对于任意实数 t, 有

$$||tx||_\infty = \max_i |tx_i| = |t| \max_i |x_i| = |t|||x||_\infty$$

即无穷范数具有齐次性得以证明。

3. 次可加性：

$$||x + y||_\infty = \max_i |x_i + y_i|$$

$$||x||_\infty = \underset{i}{max}\,|x_i|$$
$$||y||_\infty = \underset{i}{max}\,|y_i|$$
$$\underset{i}{max}\,|x_i + y_i| \le \underset{i}{max}\,|x_i| + \underset{i}{max}\,|y_i|$$

即：

$$||x + y||_\infty \le ||x||_\infty + ||y||_\infty$$

所以无穷范数的次可加性得以证明。 □

1.3 证明向量范数均是凸函数

证明. 任取$x, y \in \mathbb{R}^n$有：

$$\theta x + (1-\theta)y \in \mathbb{R}^n, \theta \in [0,1]$$

因为

$$\begin{aligned} \|\theta x + (1-\theta)y\| &\le \|\theta x\| + \|(1-\theta)y\| \\ &= \theta\,\|x\| + (1-\theta)\,\|y\| \end{aligned}$$

故向量范数均是凸函数。 □

1.4 log-sum-exp 函数的凸性

log-sum-exp 函数形式如下：$f(x) = \log(\sum_{i=1}^{k} e^{x_i})$.

证明. 由题意知

$$f(x) = \log \sum_{k=1}^{n} exp\ x_k$$

对其求一阶偏导得：

$$\frac{\partial f(x)}{\partial x_k} = \frac{e^{x_k}}{\sum_{k=1}^{n} e^{x_k}} \qquad (k = 1, 2 \cdots n)$$

对其求二阶偏导得：

当 $k \ne m$ 时，

$$\frac{\partial^2 f(x)}{\partial x_k \partial x_m} = -\frac{e^{x_k} e^{x_m}}{(\sum_{k=1}^{n} e^{x_k})^2} \qquad (k, m = 1, 2 \cdots n)$$

当 $k = m$ 时，

$$\frac{\partial^2 f(x)}{\partial x_k \partial x_m} = \frac{-e^{x_k} e^{x_m} + e^{x_k} \sum_{k=1}^{n} e^{x_k}}{(\sum_{k=1}^{n} e^{x_k})^2} \qquad (k, m = 1, 2 \cdots n)$$

下边证明该矩阵是半正定矩阵。根据半正定矩阵的定义：对任意的向量 x, 如果有 $x^T A x \ge 0$, 则为半正定矩阵。

定义 n 维列向量 $y = [e^{x_1}, e^{x_2} \cdots e^{x_n}]^T$, 则该函数的 $Hessian$ 矩阵为：

$$H(x) = \frac{1}{(y^T 1)^2}((y^T 1) diag(y) - y y^T)$$

其中 1 为全 1 向量, 因为系数为正, 所以直接证明后边的矩阵为半正定。

因为

$$x^T((y^T 1) diag(y) - y y^T)x = (y^T 1) x^T diag(y) x - x^T y y^T x = \sum_i y_i \sum_i (x_i^2 y_i) - (\sum_i x_i y_i)^2$$

令 $a_i = x_i \sqrt{y_i}, b_i = \sqrt{y_i}$, 那么由 $Cauchy - Schwarz$ 公式得：

$$(b^T b)(a^T a) - (a^T b)^2 \ge 0$$

所以该函数的 $Hessian$ 矩阵为半正定矩阵, 则该函数为凸函数。 □

1.5 如果 f 是严格凸的，那么其解集唯一，即解集 X_{opt} 只包含一个元素

证明. 假设该函数不只有一个最小值点, 其中 x 和 y($x \neq y$) 均为 f 的最小值点
根据强凸函数的等价定义, 取 $\theta \in (0,1)$, 则：

$$\begin{aligned} f(\theta x+(1-\theta)y) &\leq \theta f(x)+(1-\theta)f(y)-\frac{m}{2}\theta(1-\theta)\left\|x-y\right\|^2 \\ &= f(x)-\frac{m}{2}\theta(1-\theta)\left\|x-y\right\|^2 \\ &< f(x) \end{aligned}$$

与 f(x) 为最小值矛盾，故假设错误，原命题成立。 □

1.6 假设 f 在全空间 $\mathbb{R}^n$ 可微，如果 x^* 是一个局部极小点，那么:

$$\nabla f(x^*) = 0$$

证明. 任取 $v \in R^n$, 将 f 在点 $x = x^*$ 处泰勒展开

$$f(x^*+tv) = f(x^*) + tv^T\nabla f(x^*) + o(t)$$

整理得

$$\frac{f(x^*+tv)-f(x^*)}{t} = v^T\nabla f(x^*) + o(1)$$

根据 x^* 的最优性, 在上式中分别对 t 取点 0 处的左, 右极限可知

$$\lim_{t\to 0+}\frac{f(x^*+tv)-f(x^*)}{t} = v^T\nabla f(x^*) \geq 0$$

$$\lim_{t\to 0-}\frac{f(x^*+tv)-f(x^*)}{t} = v^T\nabla f(x^*) \leq 0$$

即对任意的 v 有 $v^T\nabla f(x^*) = 0$, 由 v 的任意性知 $\nabla f(x^*) = 0$。 □

2 第二次作业

2.1 完成强凸函数相互间等价性的证明

- f is strongly convex with parameter $m > 0$ if $g(x) = f(x) - \frac{m}{2}x^Tx$ is convex
- Jensen's inequality: Jensen's inequality for g is

$$f(\theta x+(1-\theta)y) \leq \theta f(x)+(1-\theta)y-\frac{m}{2}\theta(1-\vartheta)\left\|x-y\right\|_2^2$$

- monotonicity: monotonicity of ∇g gives

$$(\nabla f(x)-\nabla f(y))^T(x-y) \geq m\left\|x-y\right\|_2^2 \quad \forall x,y \in domf$$

  this is called strong monotonicity(covercivity) of ∇f

- secong-order condition: $\nabla^2 f \succeq mI$ for all x $\in domf$

证明. 采用循环论证的方法。首先由强凸性的定义式，已知：

$$g(x) = f(x) - \frac{m}{2}x^T x$$

为凸函数，则 $f(x)$ 为强凸函数。证明 Jensen's inequality、单调性、二阶条件都是强凸函数的相互等价条件。

(1) **Jensen's inequality**

$g(x)$ 为凸函数，$g(x)$ 必定满足凸性。由凸性性质，也是凸性的一个充要条件：

$$\exists\ \theta \in [0,1],\ x, y \in dom g,\ 使得\ g(\theta x + (1-\theta)y) \le \theta g(x) + (1-\theta)g(y)$$

带入 $g(x) = f(x) - \frac{m}{2}x^T x$，可得：

$$\begin{aligned} f(\theta x + (1-\theta)y) - \frac{m}{2}(\theta x + (1-\theta)y)^T(\theta x + (1-\theta)y) \\ \le \theta f(x) - \frac{m}{2}\theta x^T x + (1-\theta) f(y) - \frac{m}{2}(1-\theta) y^T y \end{aligned}$$

等价于：

$$\begin{aligned} f(\theta x + (1-\theta) y) \le \theta f(x) + (1-\theta) f(y) \\ + \frac{m}{2}\left[\|\theta x + (1-\theta) y\|_2^2 - \theta x^T x - (1-\theta) y^T y\right] \end{aligned}$$

等价于：

$$\begin{aligned} f(\theta x + (1-\theta) y) \le \theta f(x) + (1-\theta) f(y) \\ - \theta (1-\theta) \frac{m}{2}\|x - y\|_2^2 \end{aligned}$$

故 Jensen's inequality 与强凸性定义式互为等价条件。

(2) **单调性（monotonicity）**

由强凸性定义式可知：$g(x)$ 为凸函数 $\Longleftrightarrow$ $\nabla g(x)$ 是一个单调算子，也即：

$$(\nabla g(y) - \nabla g(x))^T (y - x) \ge 0, \forall\ x, y \in\ dom g$$

$\Longleftrightarrow$

$$(\nabla f(y) - my - \nabla f(x) + mx)^T (y - x) \ge 0, \forall\ x, y \in\ dom f$$

$\Longleftrightarrow$

$$(\nabla f(y) - \nabla f(x))^T (y - x) \ge m\|x - y\|_2^2, \forall\ x, y \in\ dom f$$

故单调性与强凸性定义式互为等价条件。

(3) **二阶条件（second-order condition）** 由强凸性定义式可知：$g(x)$ 为凸函数 $\Longleftrightarrow$ $\nabla^2 g(x)$ 半正定（**判断凸性的二阶条件**），也即：

$$\nabla^2 f(x) \succeq 0,\ \forall\ x \in dom f$$

由 $g(x) = f(x) - \frac{m}{2}x^T x$，求 $g(x)$ 的 Hessian 矩阵得：

$$\nabla^2 g(x) = \nabla^2 f(x) - mI \succeq 0$$

$\Longleftrightarrow$ $\nabla^2 f(x) \succeq mI$, 故强凸性的二阶条件与强凸性定义式互为等价条件。

□

2.2 证明ℓ_2 范数的对偶范数为ℓ_2 范数

证明. 由 Cachy-Schwartz 不等式:

$$u^T v \leq \|u\|_2 \|v\|_2$$

当且仅当 u 和 v 线性相关时，等号成立，取 $v = \dfrac{u}{\|u\|_2}$ 不等式的最大值为 $\|u\|_2$

$$\begin{aligned}\|u\|_* &= \sup\{u^T z \mid \|z\|_2 \leq 1\} \\ &= \|u\|_2\end{aligned}$$

故 ℓ_2 范数的对偶范数是 ℓ_2 范数得证。 □

2.3 完成如下邻近算子的证明

1. 二次函数

$$h(x) = \frac{1}{2}x^T Ax + b^T x + c，\operatorname{prox}_{th}(x) = (I + tA)^{-1}(x - tb)$$

2. 负自然对数的和

$$h(x) = -\sum_{i=1}^{n} \ln x_i，\operatorname{prox}_{th}(x)_i = \frac{x_i + \sqrt{x_i^2 + 4t}}{2}，i = 1, 2, ..., n$$

证明. 1. 令 $u = \operatorname{prox}_{th}(x)$, 由邻近算子的最优性条件得:

$$x - u \in \partial h(u) = tAu + tb$$

$$x - u = tAu + tb$$

化简, 得:

$$u = (I + tA)^{-1}(x - tb)$$

所以

$$\operatorname{prox}_{th}(x) = (I + tA)^{-1}(x - tb)$$

2. 令 $u_i = prox_{th}(x)_i, i = 1, 2, \cdots, n$, 由邻近算子的最优性条件得:

$$x_i - u_i \in \partial h(u_i) = -\frac{t}{u_i}(u_i > 0)$$

$$x_i - u_i = -\frac{t}{u_i}$$

解得:

$$u_i = \frac{x_i + \sqrt{x_i^2 + 4t}}{2}(u_i = \frac{x_i - \sqrt{x_i^2 + 4t}}{2} < 0\text{舍去})$$

所以

$$\operatorname{prox}_{th}(x_i) = \frac{x_i + \sqrt{x_i^2 + 4t}}{2}, i = 1, 2, \cdots, n$$

□

2.4 近似点算子基本运算规则证明

完成以下近似点算子基本运算规则的证明。

2.4.1 Affine Transformation

若 $f(x) = ag(x) + b$，其中 $a > 0$，有：$\mathrm{prox}_f(x) = \mathrm{prox}_{ag}(x)$

证明.

$$\begin{aligned}\mathrm{prox}_f(x) &= \underset{z}{\mathrm{argmin}}\{\frac{1}{2}\|z-x\|_2^2 + ag(z) + b\} \\ &= \underset{z}{\mathrm{argmin}}\{\frac{1}{2}\|z-x\|_2^2 + ag(z)\} \\ &= \mathrm{prox}_{ag}(x)\end{aligned}$$

该运算规则得以证明。 □

2.4.2 Affine Addition

若 $f(x) = g(x) + a^T x + b$，有：$\mathrm{prox}_f(x) = \mathrm{prox}_g(x - a)$

证明.

$$\begin{aligned}\mathrm{prox}_f(x) &= \underset{z}{\mathrm{argmin}}\{\frac{1}{2}\|z-x\|_2^2 + g(z) + a^T z + b\} \\ &= \underset{z}{\mathrm{argmin}}\{\frac{1}{2}(\|z\|_2^2 + \|x\|_2^2 - 2\langle z, x\rangle) + g(z) + \langle z, a\rangle + b\} \\ &= \underset{z}{\mathrm{argmin}}\{\frac{1}{2}(\|z\|_2^2 + \|x\|_2^2 - 2\langle z, x-a\rangle) + g(z)\} \\ &= \underset{z}{\mathrm{argmin}}\{\frac{1}{2}(\|z\|_2^2 + \|x\|_2^2 - 2\langle z, x-a\rangle) + g(z) + \frac{1}{2}\|a\|_2^2 - \langle x, a\rangle\} \\ &= \underset{z}{\mathrm{argmin}}\{\frac{1}{2}(\|z\|_2^2 + \|a\|_2^2 + \|x\|_2^2 - 2\langle z, x-a\rangle - 2\langle x, a\rangle) + g(z)\} \\ &= \underset{z}{\mathrm{argmin}}\{\frac{1}{2}\|z-(x-a)\|_2^2 + g(z)\} \\ &= \mathrm{prox}_g(x-a)\end{aligned}$$

该运算规则得以证明。 □

2.4.3 Quadratic Addition

若 $f(x) = g(x) + \frac{\rho}{2}\|x-a\|_2^2$，有：$\mathrm{prox}_f(x) = \mathrm{prox}_{\frac{1}{1+\rho}g}(\frac{1}{1+\rho}x + \frac{1}{1+\rho}a)$

证明.

$$
\begin{aligned}
\operatorname{prox}_f(x) &= \operatorname*{argmin}_z\{\frac{1}{2}\|z-x\|_2^2 + g(z) + \frac{\rho}{2}\|z-a\|_2^2\} \\
&= \operatorname*{argmin}_z\{\frac{1}{2}(\|z\|_2^2 + \|x\|_2^2 - 2\langle z, x\rangle) + g(z) + \frac{\rho}{2}(\|z\|_2^2 + \|a\|_2^2 - 2\langle z, a\rangle)\} \\
&= \operatorname*{argmin}_z\{\frac{1+\rho}{2}\|z\|_2^2 - \langle z, x+\rho a\rangle + g(z)\} \\
&= \operatorname*{argmin}_z\{\frac{1}{2}\|z\|_2^2 - \frac{1}{1+\rho}\langle z, x+\rho a\rangle + \frac{1}{1+\rho}g(z)\} \\
&= \operatorname*{argmin}_z\{\frac{1}{2}\left\|z - (\frac{1}{1+\rho}x + \frac{\rho}{1+\rho}a)\right\|_2^2 + \frac{1}{1+\rho}g(z)\} \\
&= \operatorname{prox}_{\frac{1}{1+\rho}g}(\frac{1}{1+\rho}x + \frac{\rho}{1+\rho}a) \qquad (1)
\end{aligned}
$$

该运算规则以证明得。 □

2.4.4 Scaling And Translation

若 $f(x) = g(ax+b)$，其中 $a \neq 0$，则有：$\operatorname{prox}_f(x) = \frac{1}{a}(\operatorname{prox}_{a^2g}(ax+b) - b)$

证明. 令 $z = \operatorname{prox}_f(x)$, 由此可知：

$$
\begin{gathered}
x - z \in \partial f(z) = \partial g(az+b) \\
(ax+b) - (az+b) \in \partial a^2 g(az+b) \\
\therefore az + b = \operatorname{prox}_{a^2g}(ax+b) \\
z = \frac{1}{a}(\operatorname{prox}_{a^2g}(ax+b) - b)
\end{gathered}
$$

将 $z = \operatorname{prox}_f(x)$ 带入上式, 可得：

$$
\operatorname{prox}_f(x) = \frac{1}{a}(\operatorname{prox}_{a^2g}(ax+b) - b)
$$

该运算规则以证明得。 □

2.5 对所有已经学过的迭代算法的收敛速度进行整理

方法	收敛步数复杂度
梯度法 (凸可微)	$O(\frac{1}{\varepsilon})$
梯度法 (强凸可微)	$O(log\frac{1}{\varepsilon})$
次梯度法 (凸不可微)	$O(\frac{1}{\varepsilon^2})$
次梯度法 (强凸不可微)	$O(\frac{1}{\varepsilon})$
邻近点梯度法 (凸不可微)	$O(\frac{1}{\varepsilon})$
邻近点梯度法 (强凸不可微)	$O(log\frac{1}{\varepsilon})$

2.6 共轭函数的凸性证明

函数 f 的共轭函数为：

$$
f^*(y) = \sup_{x \in \operatorname{dom} f}(y^T x - f(x))
$$

证明共轭函数总是为一凸函数。

方法一. 由题意知, 共轭函数的表达式如下:

$$f^*(y)=\sup_{x\in\mathrm{dom}\,f}(y^Tx-f(x))$$

设 $y_1,y_2\in\mathrm{dom}f^*,t\in[0,1]$, 将 $y=ty_1+(1-t)y_2$ 带入共轭函数中, 得:

$$\begin{aligned}
& f^*(ty_1+(1-t)y_2)\\
=& \sup_{x\in\mathrm{dom}\,f}\{(ty_1+(1-t)y_2)^Tx-f(x)\}\\
=& \sup_{x\in\mathrm{dom}\,f}\{ty_1^Tx+(1-t)y_2^Tx-(1-t)f(x)-tf(x)\}\\
=& \sup_{x\in\mathrm{dom}\,f}\{t(y_1^Tx-f(x))+(1-t)(y_2^Tx-f(x))\}\\
\le& t\sup_{x\in\mathrm{dom}\,f}\{(y_1^Tx-f(x))\}+(1-t)\sup_{x\in\mathrm{dom}\,f}\{(y_2^Tx-f(x))\}\\
=& tf^*(y_1)+(1-t)f^*(y_2)
\end{aligned}$$

综上可知, 可得到如下的关系式:

$$f^*(ty_1+(1-t)y_2)\le tf^*(y_1)+(1-t)f^*(y_2)$$

根据凸函数的性质可以判定共轭函数 $f^*(y)$ 是凸函数, 且与 $f(x)$ 的凹凸无关。 □

方法二. 由题意知，共轭函数的表达式如下:

$$f^*(y)=\sup_{x\in\mathrm{dom}\,f}(y^Tx-f(x))$$

$\because f^*(y)$ 是关于 y 的仿射函数, 且由相关证明可知, 仿射函数是凸函数。对于每个 x, 都有对应的 $(y^Tx-f(x))$ 仿射函数。

$\therefore f^*(y)$ 为一系列凸函数的逐点上确界, 由保凸运算可知,$f^*(y)$ 为凸函数。 □

2.7 ℓ_2 范数的共轭函数证明

求证 $f_0=\|x\|_2$ 的共轭函数为:

$$f_0^*(y)=\begin{cases}0 & \|y\|_*\le 1\\ \infty & otherwise\end{cases}$$

证明. 根据共轭函数定义，有:

$$f_0^*(y)=\sup_{x\in\mathrm{dom}\,f}y^Tx-f_0(x)=\sup_{x\in\mathrm{dom}\,f}y^Tx-\|x\|$$

根据对偶范数的定义，有:

$$\|y\|_*=\sup\left\{y^Tz\mid\|z\|\le 1\right\}$$

（1）当 $\|y\|_*>1$ 时，存在 $\|z\|\le 1$，使 $y^Tz>1$，故有:

$$y^Tx-\|x\|=\lambda(y^Tz-\|z\|)$$

其中，$y^Tz-\|z\|>0$。令 $\lambda\to\infty$，则有:

$$f_0^*(y)=\sup_{x\in\mathrm{dom}\,f}y^Tx-\|x\|=\sup\lambda(y^Tz-\|z\|)\to\infty$$

（2）当 $\|y\|_* \le 1$ 时，根据 Cauchy-Schwarz 不等式，有：

$$y^T x \le \|y\|_* \|x\| \le \|x\|$$

故 $y^T x - \|x\| \le 0$，即：

$$f_0^*(y) = \sup_{x \in \mathrm{dom}\, f} y^T x - \|x\| = 0$$

综上，一般范数 $\|\cdot\|$ 的共轭函数为：

$$f_0^*(y) = \begin{cases} 0 & \|y\|_* \le 1 \\ \infty & otherwise \end{cases}$$

□

2.8 Lagrange 对偶问题的凸性证明

证明：Lagrange 对偶问题总是为一凸优化问题：

$$\begin{aligned} \max \quad & g(\lambda, \nu) \\ s.t. \quad & \lambda \succeq 0 \end{aligned}$$

证明. 由于对偶问题需求解 Lagrange 对偶函数的最大值, 故要证明 Lagrange 对偶问题总是为一凸优化问题, 只需证 $g(\lambda, \upsilon)$ 为凹函数。

原问题的一般形式为：

$$\begin{aligned} & \min f(x) \\ s.t. \quad & Ax \le b \\ & Cx = d \end{aligned}$$

则其 Lagrange 对偶函数为:

$$\begin{aligned} g(\lambda, \upsilon) &= \inf_{x \in domf} (f(x) + (A^T\lambda + C^T\upsilon)^T x - b^T\lambda - d^T\upsilon) \\ &= \min[L(x_1, \lambda, \upsilon), L(x_2, \lambda, \upsilon), \cdots, L(x_n, \lambda, \upsilon)] \end{aligned} \tag{1}$$

$g(\lambda, \upsilon)$ **为凹函数的具体证明如下：**

记 $t = (\lambda, \upsilon)$，故 $g(\theta\lambda_1 + (1-\theta)\lambda_2, \theta\upsilon_1 + (1-\theta)\upsilon_2)$ 可简便记为 $g(\theta t_1 + (1-\theta)t_2)$

由 (1) 式可知：

$$\begin{cases} g(\lambda_1, \upsilon_1) = g(t_1) = \min[L(x_1, t_1), L(x_2, t_1), \cdots, L(x_n, t_1)] \\ g(\lambda_2, \upsilon_2) = g(t_2) = \min[L(x_1, t_2), L(x_2, t_2), \cdots, L(x_n, t_2)] \end{cases}$$

$$\begin{aligned} g(\theta t_1 + (1-\theta)t_2) &= \min[L(x_1, \theta t_1 + (1-\theta)t_2), \cdots, L(x_n, \theta t_1 + (1-\theta)t_2)] \\ &\ge \min[\theta L(x_1, t_1) + (1-\theta)L(x_1, t_2), \cdots, \theta L(x_n, t_1) + (1-\theta)L(x_n, t_2)] \\ &\ge \theta \min[L(x_1, t_1), \cdots, L(x_n, t_1)] + (1-\theta)\min[L(x_1, t_2), \cdots, L(x_n, t_2)] \\ &= \theta g(t_1) + (1-\theta)g(t_2) \end{aligned}$$

$$\Downarrow$$

$$g(\theta\lambda_1 + (1-\theta)\lambda_2, \theta\upsilon_1 + (1-\theta)\upsilon_2) \ge \theta g(\lambda_1, \upsilon_1) + (1-\theta)g(\lambda_2, \upsilon_2)$$

即 Lagrange 对偶函数 $g(\lambda, \upsilon)$ 满足凹函数的定义, 为凹函数。故 Lagrange 对偶问题总为凸优化问题。 □

2.9 写出岭回归的对偶问题

$$\min_x \tfrac{1}{2}\|Ax-b\|_2^2+\mu\|x\|_2^2$$

证明. 令 $Ax-b=r$，则原问题等价为

$$\min_{x,r} \frac{1}{2}\|r\|_2^2+\mu\|x\|_2^2$$
$$s.t. \quad Ax-b=r$$

其拉格朗日函数为：

$$\begin{aligned}L(x,r,v)&=\frac{1}{2}\|r\|_2^2+\mu\|x\|_2^2-<v,Ax-b-r>\\&=\frac{1}{2}\|r\|_2^2+v^Tr+\mu\|x\|_2^2-(A^Tv)^Tx+b^Tv\end{aligned}$$

由二次函数最小值的性质可得：

$$\begin{aligned}g(v)&=\inf_{x,r}L(x,r,v)\\&=-\frac{1}{2}\|v\|_2^2-\frac{1}{4\mu}\|A^Tv\|_2^2+b^Tv\end{aligned}$$

故其对偶问题是：

$$\max_v \quad -\frac{1}{2}\|v\|_2^2-\frac{1}{4\mu}\|A^Tv\|_2^2+b^Tv$$

□

2.10 自和谐函数计算

2.10.1 自和谐性质一的证明

若 f 是自和谐的, $a\geq 0$, 则 af 也是自和谐的。

证明. 若 $f(x)$ 函数是自和谐函数, 则有：

$$|f^{'''}(x)|\leq 2(f^{''}(x))^{\frac{3}{2}}$$

设函数 $g(x)=af(x),a\geq 1$, 将函数 $g(x)$ 对 x 进行二阶、三阶求导得：

$$g^{''}(x)=a^2f^{''}(x)$$

$$g^{'''}(x)=a^3f^{'''}(x)$$

综合上述不等式得：

$$|g^{'''}(x)|=|a^3f^{'''}(x)|=a^3|f^{'''}(x)|\leq 2a^3(f^{''}(x))^{\frac{3}{2}}=2(a^2f^{''}(x))^{\frac{3}{2}}=2(g^{''}(x))^{\frac{3}{2}}$$

即为：

$$|g^{'''}(x)|\leq 2(g^{''}(x))^{\frac{3}{2}}$$

由此可知 $g(x)=af(x),a\geq 1$ 为自和谐函数, 该命题得以证明。 □

2.10.2 自和谐性质二的证明

自和谐函数之和仍为自和谐函数。

证明. 若 f_1, f_2 为自和谐函数, 且 $f_1, f_2 : R \to R$, 则有：

$$
\begin{aligned}
\mid f_1^{'''}(x) + f_2^{'''}(x) \mid &\leq \mid f_1^{'''}(x) \mid + \mid f_2^{'''}(x) \mid \\
&\leq 2(f_1^{''}(x)^{\frac{3}{2}} + f_2^{''}(x)^{\frac{3}{2}}) \\
&\leq 2(f_1^{''}(x) + f_2^{''}(x))^{\frac{3}{2}}
\end{aligned}
$$

其中, 最后一个不等式用到了以下不等式：
对任意 $u, v \geq 0$, 则：

$$(u^{\frac{3}{2}} + v^{\frac{3}{2}})^{\frac{2}{3}} \leq u + v$$

该命题得以证明。 □

2.10.3 自和谐性质三的证明

仿射复合：preserved under composition with affine function.

证明. 已知 $f : R^n \to R$ 是自和谐函数, 根据定义有 $\hat{f}(t) = f(x + tv)$ 是自和谐函数, 其中 $x \in dom\ f, v \in R^n$。

令 $g(x) = f(Ax + b)$, 则有：

$$\hat{g}(t) = g(x + tv) = f(A(x + tv) + b) = f(Ax + b + tAv)$$

其中, 令 $x^{'} = Ax + b, v^{'} = Av$, 易知 $x^{'} \in dom\ f, v^{'} \in R^n$, 则有：

$$\hat{g}(t) = f(x^{'} + tv^{'}) = \hat{f}(t)$$

$\hat{g}(t)$ 为自和谐函数, 因此 $g(x) = f(Ax + b)$ 为自和谐函数。所以自和谐性是仿射不变的, 命题即可得证。 □

2.10.4 示例一的证明

若函数 $f(x) = -\sum_{i=1}^{m} \log(b_i - a_i^T x)$ on $\left\{x | a_i^T x - b_i < 0\right\}$, 证明该函数是自和谐函数。

证明. 已知 $f(x) = -\log(x)$ 是自和谐的, 由于自和谐函数的仿射不变性, 得 $f(x) = \log(b_i - a_i^T x)$ 也是自和谐的。

又自和谐对加法保持不变, 故

$$f(x) = -\sum_{i=1}^{m} \log(b_i - a_i^T x)$$

是自和谐函数。 □

2.10.5 示例二的证明

若函数 $f(x) = -\log\det X$ on S_{++}^n, 证明该函数是自和谐函数。

证明. 该函数为高维函数, 我们考虑将函数限制在直线上, 即令 $\tilde{f}(t) = f(X + tV)$, 其中 $X > 0$, $V \in S^n$. 可以将其表示为

$$\begin{aligned}
\tilde{f}(t) &= -\log\det(X^{\frac{1}{2}}(I + tX^{-\frac{1}{2}}VX^{-\frac{1}{2}})X^{\frac{1}{2}}) \\
&= -\log\det(X) - \log\det(I + tX^{-\frac{1}{2}}VX^{-\frac{1}{2}}) \\
&= -\log\det(X) - \sum_{i=1}^{n}\log(1 + t\lambda_i)
\end{aligned}$$

其中 λ_i 是 $X^{-\frac{1}{2}}VX^{-\frac{1}{2}}$ 的特征值。每个 $-\log(1 + t\lambda_i)$ 是 t 的自和谐函数, 所以他们的和是自和谐的, 由此可知 f 是自和谐的。 □

2.10.6 示例三的证明

利用对数复合规则说明以下函数是自和谐的: $f(x, y) = -\log(y^2 - x^T x)$, 定义域为 $\{(x, y) \mid \|x\|_2 < y\}$.

证明.

$$f(x, y) = -\log(y - \frac{x^T x}{y}) - \log y$$

将 f 限制在直线上, 令 $x = \hat{x} + tv$, $y = \hat{y} + tw$, 代入得

$$f(x = \hat{x} + tv, y = \hat{y} + tw) = -\log(\hat{y} + tw - \frac{\hat{x}^T\hat{x}}{\hat{y} + tw} - \frac{2t\hat{x}^T v}{\hat{y} + tw} - \frac{t^2 v^T v}{\hat{y} + tw}) - \log(\hat{y} + tw)$$

1. 当 $w = 0$ 时,

$$\text{原式} = -\log(\hat{y} - \tfrac{\hat{x}^T\hat{x}}{\hat{y}} - \tfrac{2t\hat{x}^T v}{\hat{y}} - \tfrac{t^2 v^T v}{\hat{y}}) - \log\hat{y}$$

是关于 t 的凹的二次函数的负对数与负对数的和, 这两个函数都是自和谐函数, 故该函数为自和谐函数。

2. 当 $w \neq 0$ 时, 令 $t = \frac{y - \hat{y}}{w}$ 代入, 改为关于 y 的函数, 原式可以写成

$$\begin{aligned}
&-\log(y - \frac{\hat{x}^T\hat{x}}{y} - \frac{2\frac{y-\hat{y}}{w}\hat{x}^T v}{y} - \frac{\frac{(y-\hat{y})^2}{w^2}v^T v}{y}) - \log y \\
&= -\log(-\frac{2\hat{x}^T v}{w} + \frac{2\hat{y}v^T v}{w^2} - \frac{\hat{x}^T\hat{x}}{y} + \frac{2\hat{y}\hat{x}^T v}{wy} - \frac{\hat{y}^2 v^T v}{w^2 y} + y - \frac{y v^T v}{w^2}) - \log y \\
&= -\log(\alpha + \beta y - \frac{\gamma}{y}) - \log y
\end{aligned}$$

其中

$$\begin{aligned}
\alpha &= -\frac{2\hat{x}^T v}{w} + \frac{2\hat{y}v^T v}{w^2} \\
\beta &= 1 - \frac{v^T v}{w^2} \\
\gamma &= \hat{x}^T\hat{x} - \frac{2\hat{y}\hat{x}^T v}{w} + \frac{\hat{y}^2 v^T v}{w^2}
\end{aligned}$$

定义 $g(y) = -\alpha - \beta y + \frac{\gamma}{y}$, 它是凸函数 $\gamma > 0$ 在最后证明, 且满足

$$|g^{'''}(y)| \leq \frac{3g^{''}(y)}{y},$$

其中, $g^{'''}(y) = -\frac{6\gamma}{y^4}$, $g^{''}(y) = \frac{2\gamma}{y^3}$ 故函数为自和谐函数
下面补充证明 $\gamma > 0$:

$$\begin{aligned}\gamma &= \hat{x}^T\hat{x} - \frac{2\hat{y}\hat{x}^T v}{w} + \frac{\hat{y}^2 v^T v}{w^2} \\ &= \hat{x}^2 - \frac{2\hat{y}\hat{x}v}{w} + \frac{\hat{y}^2 v^2}{w^2} \\ &= \frac{\hat{x}^2 w^2 - 2\hat{y}\hat{x}vw + \hat{y}^2 v^2}{w^2} \\ &= \frac{(\hat{x}w - \hat{y}v)^2}{w^2} \geq 0\end{aligned}$$

□

2.11 （选做）次微分基本运算规则

- 1. scaling: $\partial(\alpha f) = \alpha \partial f$ (for $\alpha > 0$)
- 2. summation: $\partial(f_1 + f_2) = \partial f_1 + \partial f_2$
- 3. affine transformation: if $h(x) = f(Ax + b), then$

$$\partial h(x) = A^T \partial f(Ax + b)$$

- 4. chain rule: suppose f is convex, and g is differentiable, nondecreasing, and convex. Let $h = g \circ f$,then

$$\partial h(x) = g'(f(x))\partial f(x)$$

证明.
- 1. 取 $g_1 \in \partial f$

$$f(y) \geq f(x) + g_1^T(y - x)$$

$$\alpha f(y) \geq \alpha f(x) + \alpha g_1^T(y - x)$$

则 $\alpha g_1 \in \partial(\alpha f)$
取 $g_2 \in \partial(\alpha f)$

$$\alpha f(y) \geq \alpha f(x) + g_2^T(y - x)$$

$$f(y) \geq f(x) + \frac{1}{\alpha} g^T(y - x)$$

则 $\frac{1}{\alpha} g_2 \in \partial f$
所以 $\partial(\alpha f) = \alpha \partial f$ (for $\alpha > 0$)

- 2. 取 $g_1 \in \partial f_1 + \partial f_2$

$$\partial f_1(y) + \partial f_2(y) \geq \partial f_1(x) + \partial f_2(x) + g_1^T(y - x)$$

$$\partial(f_1+f_2)(y) \geq \partial(f_1+f_2)(x) + g_1^T(y-x)$$

则 $g_1 \in \partial(f_1+f_2)$
取 $g_2 \in \partial(f_1+f_2)$

$$\partial(f_1+f_2)(y) \geq \partial(f_1+f_2)(x) + g_2^T(y-x)$$

$$\partial f_1(y) + \partial f_2(y) \geq \partial f_1(x) + \partial f_2(x) + g_2^T(y-x)$$

则 $g_2 \in \partial f_1 + \partial f_2$
所以 $\partial(f_1+f_2) = \partial f_1 + \partial f_2$

- 3. 取 $g_1 \in \partial f(Ax+b)$

$$\begin{aligned} f(Ay+b) &\geq f(Ax+b) + g_1^T[(Ay+b)-(Ax+b)] \\ &= f(Ax+b) + g_1^T A(y-x) \end{aligned}$$

$$h(y) \geq h(x) + (A^T g_1)^T(y-x)$$

则 $A^T g_1 \in \partial h(x)$
取 $g_2 \in \partial h(x)$

$$h(y) \geq h(x) + g_2^T(y-x)$$

$$f(Ay+b) \geq f(Ax+b) + g_2^T(y-x)$$

$$f(Ay+b) \geq f(Ax+b) + g_2^T A^{-1}[(Ay+b)-(Ax+b)](\text{仿射变换是可逆的})$$

$$f(Ay+b) \geq f(Ax+b) + ((A^{-1})^T g_2)^T[(Ay+b)-(Ax+b)]$$

则 $(A^{-1})^T g_2 = (A^T)^{-1} g_2 \in \partial f(Ax+b)$
所以 $\partial h(x) = A^T \partial f(Ax+b)$

- 4. 设 $z \in \partial g(f(x))$,$g_1 \in \partial f(x)$

$$\begin{aligned} h(x) = g(f(x)) &\geq g(f(x_0) + g_1^T(x-x_0)) \\ &\geq g(f(x_0)) + z^T g_1^T(x-x_0) \\ &= h(x_0) + (g_1 z)^T(x-x_0) \end{aligned}$$

则 $g_1 z \in \partial h(x)$
所以 $\partial h(x) = g'(f(x))\partial f(x)$

□

2.12 （选做）

- 证明 ℓ_∞ 范数的对偶范数是 ℓ_1 范数。
- 证明任意 ℓ_p 范数 $\ell_p : ||x||_p = \sqrt[p]{\sum_1^n |x_i|^p}$ 满足范数三条性质（正定性，齐次性，三角性）。
- 进一步证明任意 ℓ_p 范数的对偶范数为 ℓ_q 范数，当且仅当 $\frac{1}{p}+\frac{1}{q}=1, p \geq 1, q \geq 1$ 时成立。

证明. 1 对 $\forall u, v \in \mathbb{R}^n$

$$\|v\|_1 = \sum_{i=1}^{n} |v_i|$$

$$\|u\|_* = \sup_{v\neq 0} \frac{u^T v}{\|v\|} = \sup_{\|v\|_\infty \leq 1} u^T v$$

且对 $\forall v \in \mathbb{R}^n$ 且 $\|v\|_1 \leq 1$ 有

$$\begin{aligned} u^T v &= \sum_{i=1}^{n} u_i v_i \\ &\leq \sum_{i=1}^{n} |u_i||v_i| \end{aligned}$$

取

$$k = \{i | \max_i |u_i|\}$$

则

$$\begin{aligned} u^T v &\leq \sum_{i=1}^{n} |u_i||v_i| \\ &= \sum_{i=1}^{n} |u_k||v_i| - \sum_{i=1}^{n} (|u_k| - |u_i|)|v_i| \\ &= |u_k| \sum_{i=1}^{n} |v_i| - \sum_{i=1}^{n} (|u_k| - |u_i|)|v_i| \\ &\leq |u_k| - \sum_{i=1}^{n} (|u_k| - |u_i|)|v_i| \\ &\leq |u_k| \end{aligned}$$

即

$$u^T v \leq \|u\|_\infty$$

且当 v 满足

$$v_k = sign(u_k)$$
$$v_i = 0 \ (i \neq k)$$

时取等, 所以 $\|u\|_* = \sup_{\|v\| \leq 1} u^T v = \|u\|_\infty$

即 l_1 范数的对偶范数为 l_∞ 范数

2 1. **正定性**

$\|\boldsymbol{x}\|_p = \sqrt[p]{\sum_{i=1}^{n} |x_i|^p} \geq 0$，当且仅当$\boldsymbol{x} = 0$时取等号。

对于任意的$\forall \ i = 1, 2, ..., n$，$|x_i| \geq 0$，所以有 $\sum_{i=1}^{n} |x_i|^p \geq 0$, 所以有：

$$\|\boldsymbol{x}\|_p = \sqrt[p]{\sum_{i=1}^{n} |x_i|^p} \geq 0$$

2. **齐次性**

即证 $||c\boldsymbol{x}||_p = |c|||\boldsymbol{x}||_p$。因为

$$||c\boldsymbol{x}||_p = \sqrt[p]{\sum_{i=1}^{n}|cx_i|^p} = \sqrt[p]{\sum_{i=1}^{n}|c|^p|x_i|^p} = |c|\sqrt[p]{\sum_{i=1}^{n}|x_i|^p}$$

故齐次性得证。

3. **三角性**

即证 $||\boldsymbol{x}||_p + ||\boldsymbol{y}||_p \geq ||\boldsymbol{x}+\boldsymbol{y}||_p$ 因为：

$$||\boldsymbol{x}+\boldsymbol{y}||_p = \sqrt[n]{\sum_{i=1}^{n}|x_i+y_i|^p} \leq \sqrt[n]{\sum_{i=1}^{n}(|x_i|^p+|y_i|^p)} \leq \sqrt[n]{\sum_{i=1}^{n}|x_i|^p} + \sqrt[n]{\sum_{i=1}^{n}|y_i|^p}$$

三角性得证。

3 利用 $H\ddot{o}lder's\ inequality$ 进行证明，即：

$$\sum_{k=1}^{n}|x_k\,y_k| \leq \left(\sum_{k=1}^{n}|x_k|^p\right)^{\frac{1}{p}}\left(\sum_{k=1}^{n}|y_k|^q\right)^{\frac{1}{q}} \text{ for all } (x_1,\ldots,x_n),(y_1,\ldots,y_n)\in\mathbb{R}^n \text{ or } \mathbb{C}^n.$$

对 $\forall z, x \in \mathbb{R}^n$, 且 $\|z\|_p = (\sum\limits_{i=1}^{n}|z_i|^p)^{\frac{1}{p}} \leq 1$

$$z^Tx = \sum_{i=1}^{n} z_ix_i \leq \sum_{i=1}^{n}|z_ix_i| = \sum_{i=1}^{n}|z_i||x_i|$$

对 $\forall a, b > 0, \forall\theta \in [0,1]$ 有

$$\theta a + (1-\theta)b \geq a^{\theta}b^{(1-\theta)}$$

当且仅当 $a = b$ 时, 对 $\forall\theta \in [0,1]$ 成立

令 $\theta = \frac{1}{p}, 1-\theta = \frac{1}{q}$, 则

$$\frac{a}{p} + \frac{b}{q} \geq a^{\frac{1}{p}}b^{\frac{1}{q}}$$

则对 $\forall i = 1, 2, \ldots, n$ 有

$$\frac{1}{p}\frac{|z_i|^p}{\sum\limits_{i=1}^{n}|z_i|^p} + \frac{1}{q}\frac{|x_i|^q}{\sum\limits_{i=1}^{n}|x_i|^q} \geq \frac{|z_i|}{(\sum\limits_{i=1}^{n}|z_i|^p)^{\frac{1}{p}}}\frac{|x_i|}{(\sum\limits_{i=1}^{n}|x_i|^q)^{\frac{1}{q}}}$$

不等式两边 n 项求和得

$$\frac{1}{p} + \frac{1}{q} = 1 \geq \frac{\sum\limits_{i=1}^{n}|z_i||x_i|}{(\sum\limits_{i=1}^{n}|z_i|^p)^{\frac{1}{p}}(\sum\limits_{i=1}^{n}|x_i|^q)^{\frac{1}{q}}}$$

即

$$\sum_{i=1}^{n}|z_i||x_i| \leq \|z\|_p\|x\|_q \leq \|x\|_q$$

且令

$$|z_i|^p = \frac{|x_i|^q}{\sum_{i=1}^{n} |x_i|^q}, \quad i = 1, 2, \ldots, n$$

$$sign(z_i) = sign(x_i), \quad i = 1, 2, \ldots, n$$

可取等

所以 $\|x\|_* = \sup_{\|z\|_p \leq 1} z^T x = \|x\|_q$

即 l_p 范数的对偶范数是 l_p 范数, 其中 p, q 满足 $\frac{1}{p} + \frac{1}{q} = 1$ 且 $p, q \geq 1$

$\square$