优化模型与软件工具

主讲教师：董庆兴

华中师范大学 信息管理学院
qxdong@mail.ccnu.edu.cn

2017 年 9 月 12 日

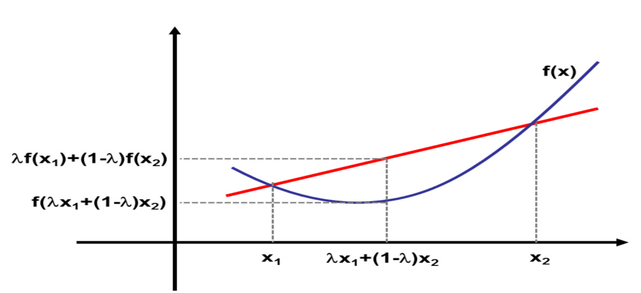
大纲

凸函数

凸函数

函数 $f:\mathbb{R}^n \to \mathbb{R}$ 是凸函数，如果 $\textbf{dom}\ f$ 是凸集，且对任何 $x, y \in \textbf{dom}\ f$ 与任何 $\theta \in [0,1]$ ，都有

$$f\left(\theta x + (1-\theta)y\right) \leq \theta f(x) + (1-\theta)f(y) \tag{1}$$



严格凸函数

严格凸函数

函数 $f:\mathbb{R}^n\to\mathbb{R}$ 是严格凸函数，如果 $\mathbf{dom}\ f$ 是凸集，且对任何 $x,y\in\mathbf{dom}\ f$ 与任何 $\theta\in[0,1]$ ，都有

$$f\left(\theta x+(1-\theta)y\right)<\theta f(x)+(1-\theta)f(y) \tag{2}$$

凸函数的图像和上境图

图像

函数 $f:\mathbb{R}^n\rightarrow\mathbb{R}$，称 $\mathbf{graph}\ f=\{(\mathbf{x},f(\mathbf{x}))|\mathbf{x}\in\mathbf{dom}\ f\}$ 为 f 的图像

上境图

函数 $f:\mathbb{R}^n\rightarrow\mathbb{R}$，称 $\mathbf{epi}\ f=\{(\mathbf{x},\mathbf{t}|\mathbf{x}\in\mathbf{dom}\ f\ f(\mathbf{x}))\leq\mathbf{t}\}$ 为 f 的上境图

- 从几何意义上看，函数 f 为凸函数，等价于 f 的上境图为凸集
- 函数 f 为凹函数，等价于 f 的亚图（上境图的对称定义）为凸集

梯度

梯度与可微

对函数 $f:\mathbb{R}^n\to\mathbb{R}$ 在 $\mathbf{x}\in\mathbf{dom}\ f$ 的某邻域内取有限值，若 f 在 $\mathbf{x}$ 处存在偏导数

$$\frac{\partial f(\mathbf{x})}{\partial x_i}=\lim_{t\to 0}\frac{f(\mathbf{x}+t\mathbf{e}_i)-f(\mathbf{x})}{t}, i=1,\cdots,n$$

其中 $\mathbf{e}_i$ 是第 i 个分量为 1，其他分量均为 0 的 n 维单位向量，并且对由 $\nabla f(\mathbf{x})=\left(\frac{\partial f(\mathbf{x})}{\partial x_1},\cdots,\frac{\partial f(\mathbf{x})}{\partial x_n}\right)^{\mathrm{T}}\in\mathbb{R}^n$ 定义的向量总有

$$f(\mathbf{x}+\mathbf{h})=f(\mathbf{x})+\nabla f(\mathbf{x})^{\mathrm{T}}\mathbf{h}+o(\|\mathbf{h}\|) \tag{3}$$

则称 f 在 $\mathbf{x}$ 处**可微**，其中 o 为高阶无穷小符号，向量 $\nabla f(\mathbf{x})$ 为 f 在 $\mathbf{x}$ 处的**梯度**

梯度常用公式

- $f(\mathbf{x}) = \mathbf{C}$，则 $\nabla f(\mathbf{x}) = \mathbf{0}$
- $f(\mathbf{x}) = \mathbf{b}^{\mathrm{T}}\mathbf{x}$，则 $\nabla f(\mathbf{x}) = \mathbf{b}$
- $f(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\mathbf{x}$，则 $\nabla f(\mathbf{x}) = 2\mathbf{x}$
- $f(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}$，且 $\mathbf{A}$ 为对称矩阵，则 $\nabla f(\mathbf{x}) = 2\mathbf{A}\mathbf{x}$

Hessian 矩阵

Hessian 矩阵

对函数 $f:\mathbb{R}^n\to\mathbb{R}$ 在 $\mathbf{x}\in\mathbf{dom}\ f$ 的某邻域内取有限值，若 f 在 $\mathbf{x}$ 处存在二阶偏导数，并且对由

$$\mathbf{H}(f)=\nabla^2 f(\mathbf{x})=\begin{pmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial x_1^2} & \frac{\partial^2 f(\mathbf{x})}{\partial x_1\partial x_2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_1\partial x_n} \\ \vdots & & & \vdots \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_n\partial x_1} & \frac{\partial^2 f(\mathbf{x})}{\partial x_n\partial x_2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_n^2} \end{pmatrix}\in\mathbb{R}^{n\times n}$$

定义的矩阵总有

$$f(\mathbf{x}+\mathbf{h})=f(\mathbf{x})+\nabla f(\mathbf{x})^{\mathrm{T}}\mathbf{h}+\frac{1}{2}\mathbf{h}^{\mathrm{T}}\nabla^2 f(\mathbf{x})\mathbf{h}+o(\|\mathbf{h}\|^2) \tag{4}$$

则称 f 在 $\mathbf{x}$ 处二阶可微，其中 o 为高阶无穷小符号，方阵 $\nabla^2 f(\mathbf{x})$ 为 f 在 $\mathbf{x}$ 处的Hessian 矩阵

Jacobi 矩阵

Jacobi 矩阵

对以 m 个函数 $F_i:\mathbb{R}^n \to \mathbb{R}(i=1,\cdots,m)$ 为分量的向量值函数 $\mathbf{F}:\mathbb{R}^n \to \mathbb{R}^m$, 定义的 $m\times n$ 矩阵

$$\mathbf{J}(\mathbf{F}) = \nabla \mathbf{F}(\mathbf{x}) = \begin{pmatrix} \nabla \mathbf{F}_1(\mathbf{x}) \\ \vdots \\ \nabla \mathbf{F}_m(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} \frac{\partial \mathbf{F}_1(\mathbf{x})}{\partial x_1} & \frac{\partial \mathbf{F}_1(\mathbf{x})}{\partial x_2} & \dots & \frac{\partial \mathbf{F}_1(\mathbf{x})}{\partial x_n} \\ \vdots & & & \vdots \\ \frac{\partial \mathbf{F}_m(\mathbf{x})}{\partial x_1} & \frac{\partial \mathbf{F}_m(\mathbf{x})}{\partial x_2} & \dots & \frac{\partial \mathbf{F}_m(\mathbf{x})}{\partial x_n} \end{pmatrix} \tag{5}$$

为 $\mathbf{F}$ 的Jacobi 矩阵

凸性判定的一阶条件

一阶凸性条件

假设 f 可微（其梯度 ∇f 在开集 $\mathbf{dom}\ f$ 内处处存在），则函数 f 是凸函数的充要条件是 $\mathbf{dom}\ f$ 是凸集且对于任意 $\mathbf{x}, \mathbf{y} \in \mathbf{dom}\ f$，下式成立：

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^{\mathrm{T}}(\mathbf{y} - \mathbf{x}) \tag{6}$$

证明：1. 必要性：由 f 是凸函数，则对于任意 $x, y \in \mathbf{dom}\ f$ 以及 $\alpha \in [0, 1]$，有

$$f[(1-\alpha)\mathbf{x} + \alpha\mathbf{y}] \leq (1-\alpha)f(\mathbf{x}) + \alpha f(\mathbf{y})$$

从而有

$$f(\mathbf{y}) - f(\mathbf{x}) \geq \frac{f[\mathbf{x} + \alpha(\mathbf{y} - \mathbf{x})] - f(\mathbf{x})}{\alpha}$$

上式两端对 $\alpha \to 0$ 求极限，右端即为 $\nabla f(\mathbf{x})^{\mathrm{T}}(\mathbf{y} - \mathbf{x})$，从而可得式 (6).

凸性判定的一阶条件

证明: 2. 充分性：选择任意 $\mathbf{x} \neq \mathbf{y} \in \mathbf{dom}\ f$ 以及 $\alpha \in [0,1]$，设 $\mathbf{z} = \alpha\mathbf{x} + (1-\alpha)\mathbf{y}$, 由式 (6) 可得

$$
\begin{aligned}
f(\mathbf{x}) &\geq f(\mathbf{z}) + \nabla f(\mathbf{z})^{\mathrm{T}}(\mathbf{x}-\mathbf{z}) \\
f(\mathbf{y}) &\geq f(\mathbf{z}) + \nabla f(\mathbf{z})^{\mathrm{T}}(\mathbf{y}-\mathbf{z})
\end{aligned}
$$

将第一个公式两端同时乘以 α，第二个式子两端同时乘以 $(1-\alpha)$，然后两端相加，可得

$$
\begin{aligned}
\alpha f(\mathbf{x}) + (1-\alpha) f(\mathbf{y}) &\geq f(\mathbf{z}) + \nabla f(\mathbf{z})^{\mathrm{T}} \times [\alpha\mathbf{x} - \alpha\mathbf{z} + (1-\alpha)(\mathbf{y}-\mathbf{z})] \\
&= f(\mathbf{z}) \\
&= f[\alpha\mathbf{x} + (1-\alpha)\mathbf{y}]
\end{aligned}
$$

命题得证。 ■

凸性判定的二阶条件

二阶凸性条件

假设 f 在 $\mathbf{dom}\ f$ 二阶连续可微，则函数 f 是凸函数的充要条件是 $\mathbf{dom}\ f$ 是凸集且对于任意 $\mathbf{x} \in \mathbf{dom}\ f$，其 Hesse 矩阵 $\mathbf{H}(\mathbf{x})$ 均为半正定

证明： 1. 充分性：由于 f 在 $\mathbf{dom}\ f$ 二阶连续可微，则对于任意 $x, y \in \mathbf{dom}\ f$ 以及 $\mathbf{H}(\mathbf{x})$ 半正定，由泰勒定理，存在 $\alpha \in (0,1)$ 有

$$f(\mathbf{x}) = f(\mathbf{y}) + \nabla f(\mathbf{y})^{\mathrm{T}}(\mathbf{x}-\mathbf{y}) + \frac{1}{2}(\mathbf{x}-\mathbf{y})^{\mathrm{T}}\mathbf{H}(\mathbf{y}+\alpha(\mathbf{x}-\mathbf{y}))(\mathbf{x}-\mathbf{y})$$

同时由凸性可知 $\mathbf{y}+\alpha(\mathbf{x}-\mathbf{y}) \in \mathbf{dom}\ f$，可知 $\mathbf{H}(\mathbf{y}+\alpha(\mathbf{x}-\mathbf{y}))$ 半正定，有 $\frac{1}{2}(\mathbf{x}-\mathbf{y})^{\mathrm{T}}\mathbf{H}(\mathbf{y}+\alpha(\mathbf{x}-\mathbf{y}))(\mathbf{x}-\mathbf{y}) \geq 0$，从而有

$$f(\mathbf{x}) \geq f(\mathbf{y}) + \nabla f(\mathbf{y})^{\mathrm{T}}(\mathbf{x}-\mathbf{y})$$

由凸函数判定一阶条件可知 f 为凸函数. ■

凸性判定的二阶条件

证明： 2. 必要性：设 $f(\mathbf{x})$ 为凸函数，对于任意 $\mathbf{x},\mathbf{z}\in\mathbf{dom}\ f$，$\mathbf{z}$ 为微小的方向向量，存在 $\bar{\alpha}>0$，使得当 $\alpha\in[-\bar{\alpha},\bar{\alpha}]$ 时，有 $\mathbf{x}+\alpha\mathbf{z}\in\mathbf{dom}\ f$，由一阶条件可得

$$f(\mathbf{x}+\alpha\mathbf{z})\geq f(\mathbf{x})+\alpha\nabla f(\mathbf{x})^{\mathrm{T}}\mathbf{z}$$

再由带 Peano 余项的泰勒公式

$$f(\mathbf{x}+\alpha\mathbf{z})=f(\mathbf{x})+\alpha\nabla f(\mathbf{x})^{\mathrm{T}}\mathbf{z}+\frac{1}{2}\alpha^2\mathbf{z}^{\mathrm{T}}\mathbf{H}(\mathbf{x})\mathbf{z}+o(\alpha^2)$$

由以上两式可得

$$\frac{1}{2}\alpha^2\mathbf{z}^{\mathrm{T}}\mathbf{H}(\mathbf{x})\mathbf{z}+o(\alpha^2)\geq 0$$

从而有

$$\frac{1}{2}\mathbf{z}^{\mathrm{T}}\mathbf{H}(\mathbf{x})\mathbf{z}+\frac{o(\alpha^2)}{\alpha^2}\geq 0$$

令 $a\to 0$，有

$$\mathbf{z}^{\mathrm{T}}\mathbf{H}(\mathbf{x})\mathbf{z}\geq 0$$

命题得证。■

凸性判定总结

- 凸函数的一阶泰勒近似总是其全局下估计
- 凸函数的函数图像具有正（向上）的曲率，也就是函数的一阶导数非减
- 凸函数上境图为凸集
- 凸函数一定弦在函数图像上方
- 任何时候都不要忘了，$\mathbf{dom}\ f$ 必须得是凸集：反例 $y = \frac{1}{x^2}$

常见凸函数: 实数集上

指数函数 对于任意 $a \in \mathbb{R}$，函数 e^{ax} 是 $\mathbb{R}$ 上的凸函数

幂函数 对于任意 $a \geq 1$ 或 $a \leq 0$ 时，函数 x^a 是正实数集 $\mathbb{R}_{++}$ 上的凸函数；当 $0 \leq a \leq 1$ 时，函数 x^a 是正实数集 $\mathbb{R}_{++}$ 上的凹函数

对数函数 函数 $\log x$ 是正实数集 $\mathbb{R}_{++}$ 上的凹函数

负熵 函数 $x \log x$ 在其定义域（$\mathbb{R}_{++}$ 或者 $\mathbb{R}_{+}$，$x = 0$ 时函数值为 0）上是凸函数

常见凸函数:$\mathbb{R}^n$ 上

范数 $\mathbb{R}^n$ 上所有范数都是凸函数：切比雪夫范数 (ℓ_p 范数)$\|\mathbf{x}\| = (|x_1|^p + \cdots + |x_n|^p)^{\frac{1}{p}}$

最大值 函数 $f(\mathbf{x}) = \max\{x_1, \cdots, x_n\}$ 在 $\mathbb{R}^n$ 上是凸的

几何平均 几何平均函数 $f(\mathbf{x}) = \left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}}$ 在定义域 $\mathbf{dom}\ f \in \mathbb{R}^n_{++}$ 上是凹函数

凸函数下水平集

凸函数下水平集是凸集

函数 $f:\mathbb{R}^n\to\mathbb{R}$ 的 $\alpha-$ 下水平集可定义为 $C_\alpha=\{\mathbf{x}\in\mathbf{dom}\ f|f(\mathbf{x})\le\alpha\}$，对于任意 α, 凸函数的下水平集为凸集

证明: 对于任意的 $\mathbf{x}^{(1)},\mathbf{x}^{(2)}\in C_\alpha$，有定义可知，有 $f(\mathbf{x}|^{(1)})\le\alpha$，$f(\mathbf{x}^{(2)})\le\alpha$. 由于 S 为凸集，因此对于每一个 $\lambda\in[0,1]$，必有

$$\lambda\mathbf{x}^{(1)}+(1-\lambda)\mathbf{x}^{(2)}\in S$$

由于 f 是定义在 S 上的凸函数，有

$$\begin{aligned}f(\lambda\mathbf{x}^{(1)}+(1-\lambda)\mathbf{x}^{(2)})&\le\lambda f(\mathbf{x}^{(1)})+(1-\lambda)f(\mathbf{x}^{(2)})\\&\le\lambda\alpha+(1-\lambda)\alpha=\alpha\end{aligned}$$

因此 $\lambda\mathbf{x}^{(1)}+(1-\lambda)\mathbf{x}^{(2)}\in C_\alpha$，故 C_α 为凸集 ■

凸函数下水平集

- 凸函数的所有下水平集一定是凸集
- 某函数的所有下水平集是凸集，不一定是凸函数：$f(\mathbf{x}) = -e^x$
- 如果 f 是凹函数，则由 $\{\mathbf{x} \in \mathbf{dom}\ f | f(\mathbf{x}) \geq \alpha\}$ 定义的 $\alpha-$ 上水平集也是凸集
- 如果某一个集合可以描述为一个凸函数的下水平集或者一个凹函数的上水平集，则该集合为凸集

Jensen 不等式

- $f(\theta \mathbf{x} + (1-\theta)\mathbf{y}) \leq \theta\, f(\mathbf{x}) + (1-\theta)\, f(\mathbf{y})$ 也称作 Jensen 不等式
- 该不等式很容易地可以延伸至 k 点的形式：如果函数 f 为凸函数，$\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_k$，$\theta_1, \theta_2, \cdots, \theta_k \geq 0$ 且 $\sum_{i=1}^{k} \theta_i = 1$，那么有

$$f(\theta_1\mathbf{x}_1 + \theta_2\mathbf{x}_2 + \cdots + \theta_k\mathbf{x}_k) \leq \theta_1 f(\mathbf{x}_1) + \theta_2 f(\mathbf{x}_2) + \cdots + \theta_k f(\mathbf{x}_k) \quad (7)$$

- 还可以延伸到无穷和与积分的形式 $\forall x \in S \subset \mathbf{dom}\ f$，$p(x) \geq 0$，且 $\int_S p(x)\mathrm{d}x = 1$，那么

$$f\left(\int_S p(x) x \mathrm{d}x\right) \leq \int_S f(x) p(x) \mathrm{d}x \quad (8)$$

- 更为一般的，我们可以在 $\mathbf{dom}\ f$ 上定义概率测度。如果 x 是定义在 $\mathbf{dom}\ f$ 上的随机变量且 f 为凸函数，那么我们有

$$f(\mathbf{E}x) \leq \mathbf{E} f(x) \quad (9)$$

- 设随机变量 x 可能取值 $\{x_1, x_2\}$，概率为 $\mathbf{prob}(x = x_1) = \theta$ $\mathbf{prob}(x = x_2) = 1-\theta$，则从上式可得 Jensen 不等式

Jensen 不等式

- 上述所有不等式都被叫做 Jensen 不等式，而实际上最初的 Jensen 不等式是
$$f(\frac{x+y}{x}) \leq \frac{f(x)+f(y)}{2}$$
- 事实上，凸性和 Jensen 不等式可以构成不等式理论的基础，考虑简单的算术-几何平均不等式 $(a, b \geq 0)$
$$\sqrt{ab} \leq \frac{a+b}{2}$$
- 我们可以利用凸性和 Jensen 不等式得到这一不等式。函数 $-\log x$ 是凸函数，利用 Jensen 不等式，令 $\theta = \frac{1}{2}$，可得
$$-\log\left(\frac{a+b}{2}\right) \leq \frac{-\log a - \log b}{2}$$
- 等式两边取指数即可得算术-几何平均不等式

非负加权求和

- 如果 f 是凸函数且 $\alpha \geq 0$，则函数 αf 也是凸函数
- 如果 f_1 和 f_2 是凸函数则函数 $f_1 + f_2$ 也是凸函数
- 那么很容易可知凸函数的非负加权求和 $f = w_1 f_1 + w_2 f_2 + \cdots + w_m f_m$ 仍是凸函数
- 同样的，凹函数的非负加权求和仍是凹函数，严格凸（凹）函数的非负，非零加权求和仍然是严格凸（凹）函数
- 该性质可以扩充到无限求和以及积分的情况

复合仿射映射

- 一个函数：$f:\mathbb{R}^n \to \mathbb{R}$ 是仿射的，如果它是一个线性函数和一个常数的和，即 $f(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{B}, \mathbf{A} \in \mathbb{R}^{m\times n}, \mathbf{b} \in \mathbb{R}^m$
- 仿射函数既是凸函数，也是凹函数
- 如果 $f:\mathbb{R}^n \to \mathbb{R}, \mathbf{A} \in \mathbb{R}^{n\times m}$ 数且 $\mathbf{b} \in \mathbb{R}^n$，则定义函数 $g:\mathbb{R}^m \to \mathbb{R}$ 为
$$g(\mathbf{x}) = f(\mathbf{A}\mathbf{x} + \mathbf{b})$$
其中 $\mathbf{dom}\ \ g = \{\mathbf{x}|\mathbf{A}\mathbf{x} + \mathbf{b} \in \mathbf{dom}\ \ f\}$
- 若 f 是凸函数，则 g 是凸函数；若 f 是凹函数，则 g 是凹函数

逐点最大的凸性

逐点最大

如果 $f_1, f_2, \cdots f_m$ 是凸函数，则二者的逐点最大值函数 f

$$f(\mathbf{x}) = \max(f_1(\mathbf{x}), f_2(\mathbf{x}), \cdots, f_m(\mathbf{x}))$$

也是凸函数

证明: 首先可知 f 定义域为
$\mathbf{dom}\ f = \mathbf{dom}\ f_1 \bigcap \mathbf{dom}\ f_2 \bigcap \cdots \mathbf{dom}\ f_m$，仍然是凸集。取 $\theta \in [0,1]$
以及 $x, y \in \mathbf{dom}\ f$，有

$$\begin{aligned}
f(\theta\mathbf{x} + (1-\theta)\mathbf{y}) &= \max\{f_1(\theta\mathbf{x} + (1-\theta)\mathbf{y}), \cdots, f_m(\theta\mathbf{x} + (1-\theta)\mathbf{y})\} \\
&\le \max\{\theta f_1(\mathbf{x}) + (1-\theta) f_1(\mathbf{y}), \cdots, \theta f_m(\mathbf{x}) + (1-\theta) f_m(\mathbf{y})\} \\
&\le \theta \max\{f_1(\mathbf{x}), \cdots, f_m(\mathbf{x})\} + (1-\theta) \max\{f_1(\mathbf{y}), \cdots, f_m(\mathbf{y})\} \\
&= \theta f(\mathbf{x}) + (1-\theta) f(\mathbf{y})
\end{aligned}$$

逐点上确界的凸性

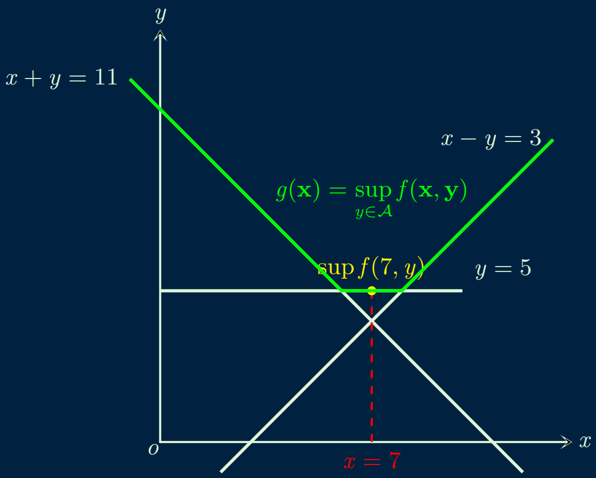
- 对于任意 $\mathbf{y}\in\mathcal{A}$，函数 $f(\mathbf{x},\mathbf{y})$ 关于 $\mathbf{x}$ 都是凸的，则函数 g

$$g(\mathbf{x})=\sup_{y\in\mathcal{A}} f(\mathbf{x},\mathbf{y}) \tag{10}$$

  关于 $\mathbf{x}$ 也是凸的，且此时 g 的定义域为
  $\mathbf{dom}\ \ g=\{\mathbf{x}|(\mathbf{x},\mathbf{y})\in\mathbf{dom}\ \ f,\forall\mathbf{y}\in\mathcal{A},\sup_{\mathbf{y}\in\mathcal{A}}f(\mathbf{x},\mathbf{y})<\infty\}$
- 同样地，我们可知一系列凹函数的逐点下确界仍然是凹函数
- 从上境图的角度理解，一系列函数的逐点上确界函数对应着这些函数上境图的交集

$$\mathbf{epi}\ \ g=\bigcap_{\mathbf{y}\in\mathcal{A}}\mathbf{epi}\ \ f(\cdot\,,\mathbf{y}) \tag{11}$$

逐点上确界的凸性


y
$x+y=11$
$x-y=3$
$g(\mathbf{x}) = \sup_{y\in\mathcal{A}} f(\mathbf{x},\mathbf{y})$
$\sup f(7,y)$
$y=5$
o
x
$x=7$

逐点上确界例子：到集合中最远点距离

- 令集合 $C \subseteq \mathbb{R}^n$，定义点 $\mathbf{x}$ 与集合 C 中最远点的距离（范数）为

$$f(\mathbf{x}) = \sup_{\mathbf{y}\in C} \|\mathbf{x} - \mathbf{y}\|$$

- 那么函数 $f(\mathbf{x})$ 是凸函数
- 对于任意 $\mathbf{y}$，函数 $\|\mathbf{x} - \mathbf{y}\|$ 关于 $\mathbf{x}$ 是凸函数。所以 f 是一族凸函数（对应不同的 $\mathbf{y} \in C$）的逐点上确界，所 f 是凸函数