凸集与凸函数

翟卫欣　副教授

zhaiweixin@cau.edu.cn

中国农业大学信息与电气工程学院

目录

目录

凸集与凸函数

- 凸集与凸函数是最优化方法理论分析中较为重要的一部分内容。
- 凸集、凸函数本身具有较好的性质。许多算法都是以凸集、凸函数为基础进行证明的。

目录

凸集的定义

- 设集合$\boldsymbol{D} \subset \mathbb{R}^n$，如果对于任意的$\boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{D}$与任意的$\alpha \in [0,1]$，有$\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y} \in \boldsymbol{D}$，则称$\boldsymbol{D}$是凸集（convex set）。

- 凸集的几何意义是：如果两个点属于此集合，则这两点连线上的任意一点均属于此集合。

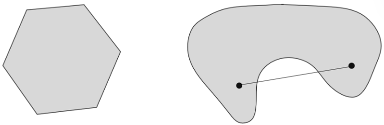

图 1: 凸集示例

凸集的性质

- 设$\boldsymbol{D}_1, \boldsymbol{D}_2 \subset \mathbb{R}^n$是凸集，$\alpha \in \mathbb{R}$，则

  (1) $\boldsymbol{D}_1 \cap \boldsymbol{D}_2 = \{\boldsymbol{x} | \boldsymbol{x} \in \boldsymbol{D}_1, \boldsymbol{x} \in \boldsymbol{D}_2\}$是凸集。

  (2) $\alpha \boldsymbol{D}_1 = \{\alpha \boldsymbol{x} | \boldsymbol{x} \in \boldsymbol{D}_1\}$是凸集。

  (3) $\boldsymbol{D}_1 + \boldsymbol{D}_2 = \{\boldsymbol{x} + \boldsymbol{y} | \boldsymbol{x} \in \boldsymbol{D}_1, \boldsymbol{y} \in \boldsymbol{D}_2\}$是凸集。

  (4) $\boldsymbol{D}_1 - \boldsymbol{D}_2 = \{\boldsymbol{x} - \boldsymbol{y} | \boldsymbol{x} \in \boldsymbol{D}_1, \boldsymbol{y} \in \boldsymbol{D}_2\}$是凸集。

凸集的性质

- 设$\boldsymbol{D}_1, \boldsymbol{D}_2 \subset \mathbb{R}^n$是凸集，则

  (1) $\boldsymbol{D}_1 \cap \boldsymbol{D}_2 = \{\boldsymbol{x} | \boldsymbol{x} \in \boldsymbol{D}_1, \boldsymbol{x} \in \boldsymbol{D}_2\}$是凸集。

- 证明：任取$\boldsymbol{x}_1 \in \boldsymbol{D}_1 \cap \boldsymbol{D}_2$、$\boldsymbol{x}_2 \in \boldsymbol{D}_1 \cap \boldsymbol{D}_2$，

  由于，$\boldsymbol{x}_1, \boldsymbol{x}_2 \in \boldsymbol{D}_1$，所以任取$\alpha' \in [0,1]$，有$\alpha' \boldsymbol{x}_1 + (1-\alpha')\boldsymbol{x}_2 \in \boldsymbol{D}_1$

  由于，$\boldsymbol{x}_1, \boldsymbol{x}_2 \in \boldsymbol{D}_2$，所以任取$\alpha' \in [0,1]$，有$\alpha' \boldsymbol{x}_1 + (1-\alpha')\boldsymbol{x}_2 \in \boldsymbol{D}_2$

  所以，任取$\boldsymbol{x}_1, \boldsymbol{x}_2 \in \boldsymbol{D}_1 \cap \boldsymbol{D}_2$，有$\alpha' \boldsymbol{x}_1 + (1-\alpha')\boldsymbol{x}_2 \in \boldsymbol{D}_1 \cap \boldsymbol{D}_2$

  所以，$\boldsymbol{D}_1 \cap \boldsymbol{D}_2$是凸集。

凸集的性质

- 设$\boldsymbol{D}_1, \boldsymbol{D}_2 \subset \mathbb{R}^n$是凸集，$\alpha \in \mathbb{R}$，则
(2) $\alpha \boldsymbol{D}_1 = \{\alpha \boldsymbol{x} | \boldsymbol{x} \in \boldsymbol{D}_1\}$是凸集。

- 证明：$\alpha = 0$时，$\alpha \boldsymbol{D}_1$是凸集。
$\alpha \neq 0$时，对任意$\boldsymbol{x}_1, \boldsymbol{x}_2 \in \boldsymbol{D}_1$，任取$\alpha' \in [0,1]$，有$\alpha' \boldsymbol{x}_1 + (1-\alpha')\boldsymbol{x}_2 \in \boldsymbol{D}_1$
由于，$\alpha \boldsymbol{D}_1$中的任一元素都可写成$\alpha \boldsymbol{x}, \boldsymbol{x} \in \boldsymbol{D}_1$的形式，
所以，任取$\alpha \boldsymbol{x}_1, \alpha \boldsymbol{x}_2 \in \alpha \boldsymbol{D}_1$，任取$\alpha' \in [0,1]$，都有$\alpha' \alpha \boldsymbol{x}_1 + (1-\alpha')\alpha \boldsymbol{x}_2 = \alpha(\alpha' \boldsymbol{x}_1 + (1-\alpha')\boldsymbol{x}_2)$，
由于，$\alpha' \boldsymbol{x}_1 + (1-\alpha')\boldsymbol{x}_2 \in \boldsymbol{D}_1$，所以$\alpha(\alpha' \boldsymbol{x}_1 + (1-\alpha')\boldsymbol{x}_2) \in \alpha \boldsymbol{D}_1$
所以，$\alpha \boldsymbol{D}_1$是凸集。

凸集的性质

- $\boldsymbol{D}_1, \boldsymbol{D}_2 \subset \mathbb{R}^n$是凸集，则
  (3) $\boldsymbol{D}_1 + \boldsymbol{D}_2 = \{\boldsymbol{x} + \boldsymbol{y} | \boldsymbol{x} \in \boldsymbol{D}_1, \boldsymbol{y} \in \boldsymbol{D}_2\}$是凸集。

- 证明：任取$\boldsymbol{x}_1, \boldsymbol{x}_2 \in \boldsymbol{D}_1$、$\boldsymbol{y}_1, \boldsymbol{y}_2 \in \boldsymbol{D}_2$，
  由于，$\boldsymbol{x}_1, \boldsymbol{x}_2 \in \boldsymbol{D}_1$，所以任取$\alpha' \in [0,1]$，有$\alpha'\boldsymbol{x}_1 + (1-\alpha')\boldsymbol{x}_2 \in \boldsymbol{D}_1$
  由于，$\boldsymbol{y}_1, \boldsymbol{y}_2 \in \boldsymbol{D}_2$，所以任取$\alpha' \in [0,1]$，有$\alpha'\boldsymbol{y}_1 + (1-\alpha')\boldsymbol{y}_2 \in \boldsymbol{D}_2$
  对任意$\boldsymbol{x}_1 + \boldsymbol{y}_1, \boldsymbol{x}_2 + \boldsymbol{y}_2 \in \boldsymbol{D}_1 + \boldsymbol{D}_2$，任取$\alpha' \in [0,1]$，有$\alpha'(\boldsymbol{x}_1 + \boldsymbol{y}_1) + (1-\alpha')(\boldsymbol{x}_2 + \boldsymbol{y}_2) = \alpha'\boldsymbol{x}_1 + (1-\alpha')\boldsymbol{x}_2 + \alpha'\boldsymbol{y}_1 + (1-\alpha')\boldsymbol{y}_2$，其中，
  $\alpha'\boldsymbol{x}_1 + (1-\alpha')\boldsymbol{x}_2 \in \boldsymbol{D}_1$，$\alpha'\boldsymbol{y}_1 + (1-\alpha')\boldsymbol{y}_2 \in \boldsymbol{D}_2$，
  所以，$\alpha'(\boldsymbol{x}_1 + \boldsymbol{y}_1) + (1-\alpha')(\boldsymbol{x}_2 + \boldsymbol{y}_2) \in \boldsymbol{D}_1 + \boldsymbol{D}_2$，
  所以，$\boldsymbol{D}_1 + \boldsymbol{D}_2$是凸集。

凸集的性质

- 设$\boldsymbol{D}_1,\boldsymbol{D}_2 \subset \mathbb{R}^n$是凸集，则
  (4) $\boldsymbol{D}_1-\boldsymbol{D}_2=\{\boldsymbol{x}-\boldsymbol{y}|\boldsymbol{x}\in\boldsymbol{D}_1,\boldsymbol{y}\in\boldsymbol{D}_2\}$是凸集。

- 证明：由凸集的性质（2）、（3）可直接推导得（4）。

凸集定义的推广

- $\boldsymbol{D}$是凸集的充分必要条件是：对任意$m \geqslant 2$，任意给定$\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_m \in \boldsymbol{D}$和实数$\alpha_1, \alpha_2, \cdots, \alpha_m$，且$\alpha_i \geqslant 0(i = 1, 2, \cdots, m; \sum_{i=1}^{m} \alpha_i = 1)$，均有 $\alpha_1\boldsymbol{x}_1 + \alpha_2\boldsymbol{x}_2 + \cdots + \alpha_m\boldsymbol{x}_m \in \boldsymbol{D}$

- 证明：当$m = 2$时，由凸集的定义，命题显然成立；假设$m = k$时命题成立，即任取$\boldsymbol{x}_i \in \boldsymbol{D}, \alpha_i \geqslant 0(i = 1, 2, \cdots, k; \sum_{i=1}^{k} \alpha_i = 1)$，则有$\sum_{i=1}^{k} \alpha_i\boldsymbol{x}_i \in \boldsymbol{D}$。

中国农业大学 China Agricultural University

凸集定义的推广

当$m=k+1$时，任取$\boldsymbol{x}_i \in \boldsymbol{D}, \alpha_i \geqslant 0(i=1,2,\cdots,k,k+1; \sum\limits_{i=1}^{k+1} \alpha_i = 1)$，则有：

$$
\begin{aligned}
\sum_{i=1}^{k+1} \alpha_i \boldsymbol{x}_i &= \sum_{i=1}^{k} \alpha_i \boldsymbol{x}_i + \alpha_{k+1} \boldsymbol{x}_{k+1} \\
&= (\sum_{i=1}^{k} \alpha_i)[\sum_{i=1}^{k} \frac{\alpha_i}{\sum\limits_{i=1}^{k} \alpha_i} \boldsymbol{x}_i] + \alpha_{k+1} \boldsymbol{x}_{k+1}
\end{aligned}
$$

由于$\sum\limits_{i=1}^{k} \frac{\alpha_i}{\sum\limits_{i=1}^{k} \alpha_i} = 1$，且$\frac{\alpha_i}{\sum\limits_{i=1}^{k} \alpha_i} \geqslant 0$，因此由归纳法则有：$\sum\limits_{i=1}^{k} \frac{\alpha_i}{\sum\limits_{i=1}^{k} \alpha_i} \boldsymbol{x}_i \in \boldsymbol{D}$

由于$\sum\limits_{i=1}^{k} \alpha_i + \alpha_{k+1} = 1$，所以$\sum\limits_{i=1}^{k+1} \alpha_i \boldsymbol{x}_i \in \boldsymbol{D}$。

开集、闭集、有界集、无界集

- 开集：对集合内任意一点，沿任意一个方向移动一个足够短的距离，仍然在集合内；
- 闭集：对集合外任意一点，沿任意一个方向移动一个足够短的距离，仍然在集合外；
- 有界集：对集合内任意一点，沿任意一个方向移动足够长的距离，肯定在集合外；
- 无界集：对集合内任意一点，存在某一个方向，沿该方向移动任意长的距离，仍然在集合内。

开集、闭集、有界集、无界集

集合形式($\mathbb{R}^n$中	开集	闭集	有界集	无界集
$\{x \mid 0 < x < 1\}$	是	否	是	否
$\{x \mid 0 \leqslant x \leqslant 1\}$	否	是	是	否
$\{x \mid 0 < x \leqslant 1\}$	否	否	是	否
$\{x \mid x > 0\}$	是	否	否	是
$\{x \mid x \geqslant 0\}$	否	是	否	是
{空集}	是	是	是	否
$\{\mathbb{R}^n\}$	是	是	否	是
$\{(x,y) \mid$当$x > 0$且$y > 0$时，满足 $x^2 + y^2 < 1$;否则，满足 $x^2 + y^2 \leqslant 1\}$	否	否	是	否

开集、闭集、有界集、无界集

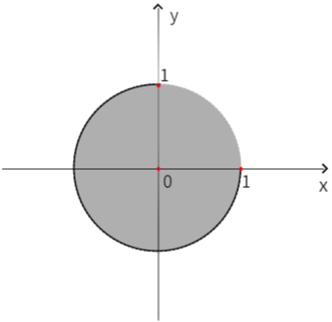

图 2: 一个既非开集又非闭集的例子

$$\{(x,y)|\text{当}x>0\text{且}y>0\text{时，满足 }x^2+y^2<1;\text{否则，满足 }x^2+y^2\leqslant 1\}$$

内点、边界、闭包

- 给定$\boldsymbol{D} \subset \mathbb{R}^n, \boldsymbol{x} \in \mathbb{R}^n$。若存在$\boldsymbol{x}$的$\delta$邻域$N_\delta(\boldsymbol{x}) = \{\boldsymbol{y} | \ ||\boldsymbol{y} - \boldsymbol{x}|| < \delta\} \subset \boldsymbol{D}$，则称$\boldsymbol{x}$为$\boldsymbol{D}$的内点；所有内点组合成的集合记为$int\boldsymbol{D}$。
- 若$\boldsymbol{x}$的任意δ邻域$N_\delta(\boldsymbol{x})$既包含$\boldsymbol{D}$中的点，又包含不属于$\boldsymbol{D}$的点，则称$\boldsymbol{x}$为$\boldsymbol{D}$的边界点；所有边界点组成的集合记为$\partial\boldsymbol{D}$。
- 若对任意$\delta > 0$均有$N_\delta(\boldsymbol{x}) \cap \boldsymbol{D} \neq \varnothing$，则称$\boldsymbol{x}$属于集合的闭包，记为$\boldsymbol{x} \in cl\boldsymbol{D}$。
- 根据以上定义可知，集合$\boldsymbol{D}$的闭包$cl\boldsymbol{D} = \boldsymbol{D} \cup \partial\boldsymbol{D}$，它是包含集合$\boldsymbol{D}$的最小的闭集。

目录

集合的分离

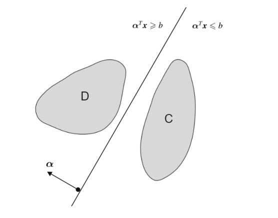

图 3: 集合的分离示例

- 设$\boldsymbol{D}_1, \boldsymbol{D}_2$是两个非空集合，$\boldsymbol{\alpha} \in \mathbb{R}^n, \boldsymbol{\beta} \in \mathbb{R}$，若有
  $\boldsymbol{D}_1 \subset \boldsymbol{H}^+ = \{\boldsymbol{x} \in \mathbb{R}^n | \boldsymbol{\alpha}^T \boldsymbol{x} \geqslant \beta\}$，
  $\boldsymbol{D}_2 \subset \boldsymbol{H}^- = \{\boldsymbol{x} \in \mathbb{R}^n | \boldsymbol{\alpha}^T \boldsymbol{x} \leqslant \beta\}$，
  则称超平面$\boldsymbol{H} = \{\boldsymbol{x} \in \mathbb{R}^n | \boldsymbol{\alpha}^T \boldsymbol{x} = \beta\}$分离集合$\boldsymbol{D}_1, \boldsymbol{D}_2$。

目录

投影定理

设$\boldsymbol{D} \subset \mathbb{R}^n$是非空闭凸集，$\boldsymbol{y} \in \mathbb{R}^n$且$\boldsymbol{y} \notin \boldsymbol{D}$，则

（1）存在唯一的一点$\bar{\boldsymbol{x}} \in \boldsymbol{D}$，使得$\bar{\boldsymbol{x}} \in \boldsymbol{D}$是$y$到$\boldsymbol{D}$的距离最小的点（距离大于0），即有$\| \bar{\boldsymbol{x}} - \boldsymbol{y} \| = min\{\| \boldsymbol{x} - \boldsymbol{y} \| \, | \boldsymbol{x} \in \boldsymbol{D}\} > 0$

（2）$\bar{\boldsymbol{x}} \in \boldsymbol{D}$是$y$到$\boldsymbol{D}$的距离最小的点的充要条件是 $(\boldsymbol{x} - \bar{\boldsymbol{x}})^T(\bar{\boldsymbol{x}} - \boldsymbol{y}) \geqslant 0, \forall \boldsymbol{x} \in \boldsymbol{D}$

（1）的证明，先证明存在：

- 设有单位球$\boldsymbol{S} = \{\boldsymbol{s} | \ \| \boldsymbol{s} \| \leqslant 1, \boldsymbol{s} \in \mathbb{R}^n\}$，取充分大的$\mu > 0$，可使$\boldsymbol{D} \cap (\boldsymbol{y} + \mu \boldsymbol{S}) \neq \varnothing$。由于$\boldsymbol{D}$是闭集，所以$\boldsymbol{D} \cap (\boldsymbol{y} + \mu \boldsymbol{S})$是非空有界闭集。
- 因此，连续函数$f(\boldsymbol{x}) = \| \boldsymbol{y} - \boldsymbol{x} \|$在$\boldsymbol{D} \cap (\boldsymbol{y} + \mu \boldsymbol{S})$上取到最小值，这个最小值在$\bar{\boldsymbol{x}} \in \boldsymbol{D} \cap (\boldsymbol{y} + \mu \boldsymbol{S})$上达到，$\bar{\boldsymbol{x}}$是$\boldsymbol{y}$到$\boldsymbol{D}$的距离最小的点。

投影定理

（1）的证明，再证明唯一性：

- 设有$\tilde{x} \in \boldsymbol{D}, \tilde{x} \neq \bar{x}$，使得$\| \tilde{x} - \boldsymbol{y} \| = \| \bar{x} - \boldsymbol{y} \| = \gamma$，则由于向量的和的模小于模的和，所以 $\| \frac{\bar{x}+\tilde{x}}{2} - \boldsymbol{y} \| \leqslant \frac{1}{2} \| \tilde{x} - \boldsymbol{y} \| + \frac{1}{2} \| \bar{x} - \boldsymbol{y} \| = \gamma$
- 由$\boldsymbol{D}$是凸集知$\frac{\bar{x}+\tilde{x}}{2} \in \boldsymbol{D}$，又因为$\gamma$是最小距离，所以上式的等号成立，即$\| \frac{\bar{x}+\tilde{x}}{2} - \boldsymbol{y} \| = \gamma$
- 所以$\bar{x} - \boldsymbol{y}$与$\tilde{x} - \boldsymbol{y}$同线且同向，即$\bar{x} - \boldsymbol{y} = \lambda(\tilde{x} - \boldsymbol{y}), \lambda > 0$，又因为已知$\| \tilde{x} - \boldsymbol{y} \| = \| \bar{x} - \boldsymbol{y} \| = \gamma$所以$\lambda = 1$，所以$\bar{x} = \tilde{x}$，唯一性得证。

投影定理

（2）的证明，先证明充分性：

- 对任意的$\boldsymbol{x} \in \boldsymbol{D}$，有

$$\begin{aligned} \| \boldsymbol{x}-\boldsymbol{y} \|^2 &= \| \boldsymbol{x}-\bar{\boldsymbol{x}}+\bar{\boldsymbol{x}}-\boldsymbol{y} \|^2 \\ &= \| \boldsymbol{x}-\bar{\boldsymbol{x}} \|^2 + \| \bar{\boldsymbol{x}}-\boldsymbol{y} \|^2 + 2(\boldsymbol{x}-\bar{\boldsymbol{x}})^T(\bar{\boldsymbol{x}}-\boldsymbol{y}) \end{aligned}$$

- 由于$\| \boldsymbol{x}-\bar{\boldsymbol{x}} \|^2 \geqslant 0$且$(\boldsymbol{x}-\bar{\boldsymbol{x}})^T(\bar{\boldsymbol{x}}-\boldsymbol{y}) \geqslant 0$，所以有：$\| \boldsymbol{x}-\boldsymbol{y} \|^2 \geqslant \| \bar{\boldsymbol{x}}-\boldsymbol{y} \|^2$，即$\bar{x}$是距离最小点。

投影定理

（2）的证明，再证明必要性：

- 由$\bar{x}$是y到D的最小点，可知$\| \bar{x} - y \| \leqslant \| x - y \|, \forall x \in D$，或等价地有：
  $\| \bar{x} - y \|^2 \leqslant \| x - y \|^2, \forall x \in D$

- 因为$\bar{x} \in D$，且D是凸集，所以对任意的$\alpha \in (0,1)$有
  $\bar{x} + \alpha(x - \bar{x}) = \alpha x + (1-\alpha)\bar{x} \in D$
  所以
  $$\begin{aligned} \| \bar{x} - y \|^2 &\leqslant \| y - \bar{x} - \alpha(x - \bar{x}) \|^2 \\ &\leqslant \| y - \bar{x} \|^2 + \alpha^2 \| x - \bar{x} \|^2 - 2\alpha(x - \bar{x})^T(y - \bar{x}) \end{aligned}$$

- 由此，可得$\alpha^2 \| x - \bar{x} \|^2 - 2\alpha(x - \bar{x})^T(y - \bar{x}) \geqslant 0, \forall \alpha \in (0,1)$

- 上式两边同除以α，并令$\alpha \to 0$，可得$(x - \bar{x})^T(\bar{x} - y) \geqslant 0, \forall \alpha \in (0,1)$

投影定理

（2）的必要性证明的补充

对于$\alpha > 0$，如果当$\alpha \to 0$时$o(\alpha) > 0$且$o(\alpha) \to 0$，则有

$$f(\boldsymbol{x}) + \frac{o(\alpha)}{\alpha} \leqslant 0, \forall \boldsymbol{x} \in \mathbb{R}^n \Rightarrow f(\boldsymbol{x}) < 0, \forall \boldsymbol{x} \in \mathbb{R}^n$$

$$f(\boldsymbol{x}) + \frac{o(\alpha)}{\alpha} \geqslant 0, \forall \boldsymbol{x} \in \mathbb{R}^n \Rightarrow f(\boldsymbol{x}) \geqslant 0, \forall \boldsymbol{x} \in \mathbb{R}^n$$

目录

点与凸集的分离定理

- 设$\boldsymbol{D} \subset \mathbb{R}^n$是非空闭凸集，$\boldsymbol{y} \in \mathbb{R}^n$，$\boldsymbol{y} \notin \boldsymbol{D}$，则存在非零向量$\boldsymbol{\alpha} \in \mathbb{R}^n, \beta \in \mathbb{R}$满足$\boldsymbol{\alpha}^T\boldsymbol{x} \leqslant \beta < \boldsymbol{\alpha}^T\boldsymbol{y}, \forall \boldsymbol{x} \in \boldsymbol{D}$

- 证明：由投影定理可知，存在唯一的最小距离点$\bar{\boldsymbol{x}} \in \boldsymbol{D}$满足
  $(\boldsymbol{x}-\bar{\boldsymbol{x}})^T(\bar{\boldsymbol{x}}-\boldsymbol{y}) \geqslant 0, \forall \boldsymbol{x} \in \boldsymbol{D}$，
  即$\boldsymbol{x}^T(\boldsymbol{y}-\bar{\boldsymbol{x}}) \leqslant \bar{\boldsymbol{x}}^T(\boldsymbol{y}-\bar{\boldsymbol{x}}), \forall \boldsymbol{x} \in \boldsymbol{D}$
  而$\| \boldsymbol{y}-\bar{\boldsymbol{x}} \|^2 = (\boldsymbol{y}-\bar{\boldsymbol{x}})^T(\boldsymbol{y}-\bar{\boldsymbol{x}}) = \boldsymbol{y}^T(\boldsymbol{y}-\bar{\boldsymbol{x}}) - \bar{\boldsymbol{x}}^T(\boldsymbol{y}-\bar{\boldsymbol{x}}), \forall \boldsymbol{x} \in \boldsymbol{D}$
- 结合以上两式，得$\| \boldsymbol{y}-\bar{\boldsymbol{x}} \|^2 \leqslant \boldsymbol{y}^T(\boldsymbol{y}-\bar{\boldsymbol{x}}) - \boldsymbol{x}^T(\boldsymbol{y}-\bar{\boldsymbol{x}}), \forall \boldsymbol{x} \in \boldsymbol{D}$
- 令$\boldsymbol{\alpha} = \boldsymbol{y}-\bar{\boldsymbol{x}}$，显然$\boldsymbol{\alpha} \neq \boldsymbol{0}$，则上式成为 $0 < \| \boldsymbol{\alpha} \|^2 \leqslant \boldsymbol{y}^T\boldsymbol{\alpha} - \boldsymbol{x}^T\boldsymbol{\alpha}, \forall \boldsymbol{x} \in \boldsymbol{D}$，则$\boldsymbol{\alpha}^T\boldsymbol{x} < \boldsymbol{\alpha}^T\boldsymbol{y}, \forall \boldsymbol{x} \in \boldsymbol{D}$
- 令$\beta = max\{\boldsymbol{\alpha}^T\boldsymbol{x}\}$，则有$\boldsymbol{\alpha}^T\boldsymbol{x} \leqslant \beta < \boldsymbol{\alpha}^T\boldsymbol{y}, \forall \boldsymbol{x} \in \boldsymbol{D}$

点与凸集的分离定理

- 推论1：设$\boldsymbol{D} \subset \mathbb{R}^n$是非空凸集，$\boldsymbol{y} \in \partial \boldsymbol{D}$，则存在非零向量$\boldsymbol{\alpha} \in \mathbb{R}^n$满足$\boldsymbol{\alpha}^{\boldsymbol{T}}\boldsymbol{x} \leqslant \boldsymbol{\alpha}^{\boldsymbol{T}}\boldsymbol{y}, \forall \boldsymbol{x} \in cl\boldsymbol{D}$

- 推论2：设$\boldsymbol{D} \subset \mathbb{R}^n$是非空凸集，$\boldsymbol{y} \notin \boldsymbol{D}$，则存在非零向量$\boldsymbol{\alpha} \in \mathbb{R}^n$满足$\boldsymbol{\alpha}^{\boldsymbol{T}}\boldsymbol{x} \leqslant \boldsymbol{\alpha}^{\boldsymbol{T}}\boldsymbol{y}, \forall \boldsymbol{x} \in cl\boldsymbol{D}$

- 注:cl是闭包，∂是边界

中国农业大学 China Agricultural University

目录

两个非空凸集的分离定理

- 设$\boldsymbol{D}_1, \boldsymbol{D}_2 \subset \mathbb{R}^n$是非空凸集，且$\boldsymbol{D}_1 \cap \boldsymbol{D}_2 = \varnothing$则存在非零向量$\boldsymbol{\alpha}$使得$inf\{\boldsymbol{\alpha}^T\boldsymbol{x}|\boldsymbol{x} \in \boldsymbol{D}_1\} \geqslant sup\{\boldsymbol{\alpha}^T\boldsymbol{x}|\boldsymbol{x} \in \boldsymbol{D}_2\}$

- 证明：令$\boldsymbol{D}^{'} = \boldsymbol{D}_2 - \boldsymbol{D}_1 = \{z|z = \boldsymbol{x}_2 - \boldsymbol{x}_1, \boldsymbol{x}_1 \in \boldsymbol{D}_1, \boldsymbol{x}_2 \in \boldsymbol{D}_2\}$，由于$\boldsymbol{D}_1, \boldsymbol{D}_2$非空，所以$\boldsymbol{D}^{'}$非空，由于$\boldsymbol{D}_1 \cap \boldsymbol{D}_2 = \varnothing$，所以$\boldsymbol{0} \notin \boldsymbol{D}^{'}$，根据点与凸集的分离定理的推论可知存在非零向量$\boldsymbol{\alpha}$，使得对每一个$z \in \boldsymbol{D}^{'}$都有$\boldsymbol{\alpha}^T z \leqslant 0$，即$\boldsymbol{\alpha}^T\boldsymbol{x}_1 \geqslant \boldsymbol{\alpha}^T\boldsymbol{x}_2, \forall \boldsymbol{x}_1 \in \boldsymbol{D}_1, \boldsymbol{x}_2 \in \boldsymbol{D}_2$

- 命题得证
- 注：sup是supremum的简写，上确界；inf是infimum的简写，下确界，$sup\{x|1 < x < 2\} = sup\{x|1 \leqslant x \leqslant 2\} = 2$

支撑超平面定理

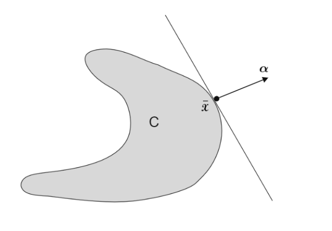

图 4: 支撑超平面示例

- 设$\boldsymbol{D} \subset \mathbb{R}^n$是非空凸集，$\bar{\boldsymbol{x}} \in \partial \boldsymbol{D}$，则存在非零向量$\boldsymbol{\alpha} \in \mathbb{R}^n$，使得$\boldsymbol{\alpha}^T\boldsymbol{x} \leqslant \boldsymbol{\alpha}^T\bar{\boldsymbol{x}}, \forall \boldsymbol{x} \in cl\boldsymbol{D}$；此时也称超平面$\boldsymbol{H} = \{\boldsymbol{x} \in \mathbb{R}^n | \boldsymbol{\alpha}^T\boldsymbol{x} = \boldsymbol{\alpha}^T\bar{\boldsymbol{x}}\}$为集合$\boldsymbol{D}$在$\bar{\boldsymbol{x}}$处的支撑超平面。

支撑超平面定理

- 证明：由于$\bar{\boldsymbol{x}} \in \partial \boldsymbol{D}$，所以存在点列$\{\boldsymbol{x}_k\}$，使得$\boldsymbol{x}_k \to \bar{\boldsymbol{x}}$，且$\boldsymbol{x}_k \notin cl\boldsymbol{D}$。根据点与凸集的分离定理可知对于每个$\boldsymbol{x}_k$，存在非零向量$\boldsymbol{\alpha}_k \in \mathbb{R}^n$，满足$\boldsymbol{\alpha}_k^T \boldsymbol{x} < \boldsymbol{\alpha}_k^T \boldsymbol{x}_k, \forall \boldsymbol{x} \in cl\boldsymbol{D}$
  不妨设$\| \boldsymbol{\alpha}_k \| = 1$，则$\{\boldsymbol{\alpha}_k\}$是有界序列，必存在收敛子列$\{\boldsymbol{\alpha}_k\}$，不妨仍记该收敛子列为$\{\boldsymbol{\alpha}_k\}$其收敛点为$\boldsymbol{\alpha}$
- 取极限可得$\boldsymbol{\alpha}^T \boldsymbol{x} \leqslant \boldsymbol{\alpha}^T \bar{\boldsymbol{x}}, \forall \boldsymbol{x} \in cl\boldsymbol{D}$。

目录

凸函数

- 设$D \subset \mathbb{R}^n$是非空凸集，$f(\boldsymbol{x})$是定义在D上的函数，如果对任意的$\boldsymbol{x}_1, \boldsymbol{x}_2 \in D, \alpha \in (0,1)$都有$f(\alpha \boldsymbol{x}_1 + (1-\alpha)\boldsymbol{x}_2) \leqslant \alpha f(\boldsymbol{x}_1) + (1-\alpha) f(\boldsymbol{x}_2)$，则称$f$是$D$上的凸函数(convex function)。
- 如果对任意的$\boldsymbol{x}_1, \boldsymbol{x}_2 \in D, \alpha \in (0,1)$都有$f(\alpha \boldsymbol{x}_1 + (1-\alpha)\boldsymbol{x}_2) < \alpha f(\boldsymbol{x}_1) + (1-\alpha) f(\boldsymbol{x}_2)$，则称$f$是$D$上的严格凸函数(strictly convex function)。
- 若f为凸函数，则$-f$为凹函数(concave function)。

凸函数举例

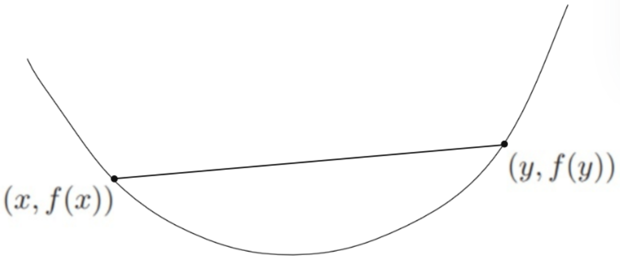

图 5: 凸函数示例

- 下列函数均为$\mathbb{R}^n$上的凸函数：$f(\boldsymbol{x}) = \boldsymbol{c}^T\boldsymbol{x}$

  $f(\boldsymbol{x}) = \| \boldsymbol{x} \|$

  $f(\boldsymbol{x}) = \boldsymbol{x}^T A \boldsymbol{x}$，其中$A$为对称正定矩阵

凸函数的α水平集

- $f(\boldsymbol{x})$是定义在$\boldsymbol{D} \subset \mathbb{R}^n$上的函数，$\alpha \in \mathbb{R}$，集合$\boldsymbol{D}_\alpha = \{\boldsymbol{x} | f(\boldsymbol{x}) \leqslant \alpha, \boldsymbol{x} \in \boldsymbol{D}\}$称作$f$函数的$\alpha$水平集(level set)。

- 求证非空凸集$\boldsymbol{D}$上凸函数f的任意水平集$\boldsymbol{D}_\alpha$是凸集。

- 证明：对任意$\boldsymbol{x}_1, \boldsymbol{x}_2 \in \boldsymbol{D}_\alpha$，有$f(\boldsymbol{x}_1) \leqslant \alpha, f(\boldsymbol{x}_2) \leqslant \alpha$，对任意的$\lambda \in (0,1)$，有$f(\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2) \leqslant \lambda f(\boldsymbol{x}_1) + (1-\lambda) f(\boldsymbol{x}_2) \leqslant \lambda\alpha + (1-\lambda)\alpha = \alpha$

- 所以$\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2 \in \boldsymbol{D}_\alpha$，所以$\boldsymbol{D}_\alpha$是凸集。

凸函数的α水平集

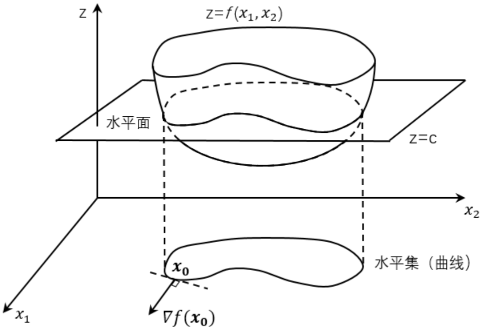

图 6: 水平集示例

图中展示了某一函数（不一定是凸函数）的水平集

目录

凸函数的判别定理1

定义在$\mathbb{R}^n$上的$f(\boldsymbol{x})$为凸函数的充要条件是对于任意$\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^n$，一元函数$\phi(\alpha)=f(\boldsymbol{x}+\alpha\boldsymbol{y})$是关于$\alpha$的凸函数。

证明：必要性

- 设$\lambda_1\geqslant 0,\lambda_2\geqslant 0,\lambda_1+\lambda_2=1$，由$\phi(\alpha)$的定义和$f(\boldsymbol{x})$的凸性，有：

$$\begin{aligned}\phi(\lambda_1\alpha_1+\lambda_2\alpha_2)&=f(\boldsymbol{x}+(\lambda_1\alpha_1+\lambda_2\alpha_2)\boldsymbol{y})\\&=f(\lambda_1\boldsymbol{x}+\lambda_2\boldsymbol{x}+(\lambda_1\alpha_1+\lambda_2\alpha_2)\boldsymbol{y})\\&=f(\lambda_1(\boldsymbol{x}+\alpha_1\boldsymbol{y})+\lambda_2(\boldsymbol{x}+\alpha_2\boldsymbol{y}))\\&\leqslant\lambda_1 f(\boldsymbol{x}+\alpha_1\boldsymbol{y})+\lambda_2 f(\boldsymbol{x}+\alpha_2\boldsymbol{y})\\&=\lambda_1\phi(\alpha_1)+\lambda_2\phi(\alpha_2)\end{aligned}$$

- 由定义知$\phi(\alpha)$是凸函数。

凸函数的判别定理1

证明：充分性

- 任取$\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^n$，记$\boldsymbol{x}=\boldsymbol{x}+0(\boldsymbol{y}-\boldsymbol{x}),\boldsymbol{y}=\boldsymbol{x}+1(\boldsymbol{y}-\boldsymbol{x}),\phi(\alpha)=f(\boldsymbol{x}+\alpha(\boldsymbol{y}-\boldsymbol{x}))$，则对于$\lambda_1\geqslant 0,\lambda_2\geqslant 0,\lambda_1+\lambda_2=1$

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

- 所以$f(\boldsymbol{x})$是凸函数。

凸函数的判别定理2

设$\boldsymbol{D} \subset \mathbb{R}^n$是非空开凸集，$f : \boldsymbol{D} \subset \mathbb{R}^n \to \mathbb{R}$，且$f(\boldsymbol{x})$在$\boldsymbol{D}$上一阶连续可微，则$f(\boldsymbol{x})$是$\boldsymbol{D}$上的凸函数的充要条件是：

$$f(\boldsymbol{y}) \geqslant f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T(\boldsymbol{y} - \boldsymbol{x}), \forall \boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{D}$$

证明：必要性

- 设$f(\boldsymbol{x})$是$\boldsymbol{D}$上的凸函数，则$\forall \alpha \in (0, 1)$，有
  $f(\alpha \boldsymbol{y} + (1 - \alpha)\boldsymbol{x}) \leqslant \alpha f(\boldsymbol{y}) + (1 - \alpha) f(\boldsymbol{x})$，
  故$\frac{f(\boldsymbol{x}+\alpha(\boldsymbol{y}-\boldsymbol{x}))-f(\boldsymbol{x})}{\alpha} \leqslant f(\boldsymbol{y}) - f(\boldsymbol{x})$
- 由Taylor展开式可知：
  $f(\boldsymbol{x} + \alpha(\boldsymbol{y} - \boldsymbol{x})) - f(\boldsymbol{x}) = \alpha \nabla f(\boldsymbol{x})^T(\boldsymbol{y} - \boldsymbol{x}) + o(\alpha \parallel \boldsymbol{y} - \boldsymbol{x} \parallel)$
  因此得到：$\nabla f(\boldsymbol{x})^T(\boldsymbol{y} - \boldsymbol{x}) + \frac{o(\alpha \|\boldsymbol{y}-\boldsymbol{x}\|)}{\alpha} \leqslant f(\boldsymbol{y}) - f(\boldsymbol{x})$
- 两边取$\alpha \to 0$的极限，有$\nabla f(\boldsymbol{x})^T(\boldsymbol{y} - \boldsymbol{x}) < f(\boldsymbol{y}) - f(\boldsymbol{x})$，比凸函数的条件更为严格（此处是$<$，凸函数只要求$\leqslant$）
- 必要性得证

凸函数的判别定理2

证明：充分性

- 设$\nabla f(\boldsymbol{x})^T(\boldsymbol{y}-\boldsymbol{x}) \leqslant f(\boldsymbol{y})-f(\boldsymbol{x}), \forall \boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{D}, \forall \alpha \in (0,1)$
  取$\bar{\boldsymbol{x}} = \alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y}$，由于$\boldsymbol{D}$是凸集，所以$\bar{\boldsymbol{x}} \in \boldsymbol{D}$
- 依据$\boldsymbol{x}, \bar{\boldsymbol{x}} \in \boldsymbol{D}; \boldsymbol{y}, \bar{\boldsymbol{x}} \in \boldsymbol{D}$，分别有：
  $f(\bar{\boldsymbol{x}}) + \nabla f(\bar{\boldsymbol{x}})^T(\boldsymbol{x}-\bar{\boldsymbol{x}}) \leqslant f(\boldsymbol{x}), \forall \boldsymbol{x} \in \boldsymbol{D}$
  $f(\bar{\boldsymbol{x}}) + \nabla f(\bar{\boldsymbol{x}})^T(\boldsymbol{y}-\bar{\boldsymbol{x}}) \leqslant f(\boldsymbol{y}), \forall \boldsymbol{y} \in \boldsymbol{D}$
- 上两式分别乘以$\alpha, (1-\alpha)$，相加得
  $f(\bar{\boldsymbol{x}}) + \nabla f(\bar{\boldsymbol{x}})^T(\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y} - \bar{\boldsymbol{x}}) \leqslant \alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{y})$
- 又因为已知$\bar{\boldsymbol{x}} = \alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y}$，得到
  $f(\bar{\boldsymbol{x}}) \leqslant \alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{y})$
  即$f(\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y}) \leqslant \alpha f(\boldsymbol{x}) + (1-\alpha) f(\boldsymbol{y})$
- 对任意的$\alpha \in (0,1)$成立，所以可知$f(\boldsymbol{x})$是凸集$\boldsymbol{D}$上的凸函数。

凸函数的判别定理2

- 设$\boldsymbol{D} \subset \mathbb{R}^n$是非空开凸集，$f : \boldsymbol{D} \subset \mathbb{R}^n \to \mathbb{R}$，且$f(\boldsymbol{x})$在$\boldsymbol{D}$上一阶连续可微，则$f(\boldsymbol{x})$是$\boldsymbol{D}$上的严格凸函数的充要条件是：
  $f(\boldsymbol{y}) > f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T(\boldsymbol{y} - \boldsymbol{x}), \forall \boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{D}$且$\boldsymbol{x} \neq \boldsymbol{y}$

凸函数的判别定理3

设$\boldsymbol{D}\subset\mathbb{R}^n$是非空开凸集，$f:\boldsymbol{D}\subset\mathbb{R}^n\to\mathbb{R}$，且$f(\boldsymbol{x})$在$\boldsymbol{D}$上二阶连续可微，则$f(\boldsymbol{x})$是$\boldsymbol{D}$上的凸函数的充要条件是：

$f(\boldsymbol{x})$的Hesse矩阵$\nabla^2 f(\boldsymbol{x})$在$\boldsymbol{D}$上是半正定的。

证明：必要性

- 任取$\bar{\boldsymbol{x}}\in\boldsymbol{D}$，由$\boldsymbol{D}$是开凸集知，$\forall\boldsymbol{x}\neq\boldsymbol{0}$，存在$\delta>0$，使得当$\alpha\in(0,\delta)$时，有$\bar{\boldsymbol{x}}+\alpha\boldsymbol{x}\in\boldsymbol{D}$，由于$f(\boldsymbol{x})$是$\boldsymbol{D}$上的凸函数，因此根据判别定理2有：
  $f(\bar{\boldsymbol{x}})+\alpha\nabla f(\bar{\boldsymbol{x}})^T\boldsymbol{x}\leqslant f(\bar{\boldsymbol{x}}+\alpha\boldsymbol{x}),\forall\boldsymbol{x}\in\boldsymbol{D}$
- 又由于$f(\boldsymbol{x})$二阶连续可微，所以按照二阶Taylor展开式有 $f(\bar{\boldsymbol{x}}+\alpha\boldsymbol{x})=f(\bar{\boldsymbol{x}})+\alpha\nabla f(\bar{\boldsymbol{x}})^T\boldsymbol{x}+\frac{1}{2}\alpha^2\boldsymbol{x}^T\nabla^2 f(\bar{\boldsymbol{x}})^T\boldsymbol{x}+o(\|\alpha\boldsymbol{x}\|^2)$
- 所以$\frac{1}{2}\alpha^2\boldsymbol{x}^T\nabla^2 f(\bar{\boldsymbol{x}})^T\boldsymbol{x}+o(\|\alpha\boldsymbol{x}\|^2)\geqslant 0,\forall\boldsymbol{x}\in\boldsymbol{D}$
- 将上式两边同除以α^2，并令$\alpha\to 0$，得$\boldsymbol{x}^T\nabla^2 f(\bar{\boldsymbol{x}})^T\boldsymbol{x}\geqslant 0,\forall\boldsymbol{x}\in\boldsymbol{D}$
- 所以$\nabla^2 f(\bar{\boldsymbol{x}})$在$\boldsymbol{D}$上是半正定的。

凸函数的判别定理3

证明：充分性

- 设$\nabla^2 f(\bar{x})$在D上是半正定的，对$\bar{x}, x \in D$将$f(x)$在$\bar{x}(\bar{x} \in D)$处作Taylor展开：
  $f(x) = f(\bar{x}) + \nabla f(\bar{x})^T(x-\bar{x}) + \frac{1}{2}(x-\bar{x})^T\nabla^2 f(\hat{x})^T(x-\bar{x})$
  $\hat{x} = \lambda x + (1-\lambda)\bar{x}, \lambda \in (0,1)$。
- 由于D是凸集，所以$\hat{x} \in D$，
- 又因为$(x-\bar{x})^T\nabla^2 f(\hat{x})^T(x-\bar{x}) \geqslant 0$，所以$f(x) \geqslant f(\bar{x}) + \nabla f(\bar{x})^T(x-\bar{x})$，
- 所以根据判别定理2有$f(x)$是D上的凸函数。

凸函数的判别定理4

- 设$\boldsymbol{D} \subset \mathbb{R}^n$是非空开凸集，$f: \boldsymbol{D} \subset \mathbb{R}^n \to \mathbb{R}$，且$f(\boldsymbol{x})$在$\boldsymbol{D}$上二阶连续可微，则$f(\boldsymbol{x})$的Hesse矩阵$\nabla^2 f(\boldsymbol{x})$在$\boldsymbol{D}$上是正定的$\Rightarrow$ $f(\boldsymbol{x})$是$\boldsymbol{D}$上的严格凸函数;
  $f(\boldsymbol{x})$是$\boldsymbol{D}$上的严格凸函数$\Rightarrow$ $f(\boldsymbol{x})$的Hesse矩阵$\nabla^2 f(\boldsymbol{x})$在$\boldsymbol{D}$上是半正定的。

- 几个反例：$y = x^4, x \in \mathbb{R}$，$y = x^6, x \in \mathbb{R}$

- 导致充分严格凸函数与严格正定不构成互为充要条件的原因：
  对于$\alpha > 0$，如果当$\alpha \to 0$时$o(\alpha) > 0$且$o(\alpha) \to 0$，则有
  $f(\boldsymbol{x}) + \frac{o(\alpha)}{\alpha} < 0, \forall \boldsymbol{x} \in \mathbb{R}^n \Rightarrow f(\boldsymbol{x}) < 0, \forall \boldsymbol{x} \in \mathbb{R}^n$
  $f(\boldsymbol{x}) + \frac{o(\alpha)}{\alpha} > 0, \forall \boldsymbol{x} \in \mathbb{R}^n \Rightarrow f(\boldsymbol{x}) \geqslant 0, \forall \boldsymbol{x} \in \mathbb{R}^n$

Thank you for your attention!

翟卫欣 副教授

zhaiweixin@cau.edu.cn

中国农业大学信息与电气工程学院