第 2 章：数学基础

(2)：保凸运算与超平面分离

授课教师：曹语

课程主页：https://yucaoyc.github.io/math3806

背景和目标

我们之前介绍了大量凸集的例子（如超平面、半空间、球、多面体等）。

目标：继续深入理解凸集的性质，并能熟练判断“这个集合是否凸”。

- **保凸运算**：如何利用简单的凸集构建复杂的凸集?
- **超平面分离定理**：凸集最核心的几何性质之一。

目录

(1) 交集运算保凸

定理 (交集运算是保凸的)

- 如果 S_1 和 S_2 是凸集，那么 $S_1 \cap S_2$ 也是凸集。
- 更一般地：如果对任意 $\alpha \in \mathcal{A}$，集合 S_α 都是凸集，则

$$\bigcap_{\alpha \in \mathcal{A}} S_\alpha$$

也是凸集（即：任意多个凸集的交集仍然是凸集）。

证明：若 $\boldsymbol{x}, \boldsymbol{y} \in \cap_\alpha S_\alpha$，则对每个 α 有 $\boldsymbol{x}, \boldsymbol{y} \in S_\alpha$，由 S_α 的凸性得 $\theta \boldsymbol{x} + (1-\theta)\boldsymbol{y} \in S_\alpha$，从而属于交集。

例子： **半正定锥 $\mathbf{S}_+^n$** 是凸集：

$$\mathbf{S}_+^n = \bigcap_{z \neq 0} \underbrace{\{\boldsymbol{X} \in \mathbf{S}^n \mid \boldsymbol{z}^\top \boldsymbol{X} \boldsymbol{z} \geqslant 0\}}_{\text{半空间（凸集）}}$$

其中 $\boldsymbol{z}^\top \boldsymbol{X} \boldsymbol{z}$ 关于 $\boldsymbol{X}$ 是线性的，所以大括号内确实是一个半空间；因此 $\mathbf{S}_+^n$ 作为无穷多个半空间的交集必为凸集。

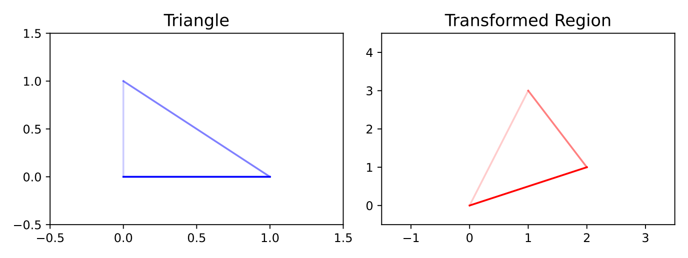
(2) 仿射函数保凸

定义 (仿射函数)

若函数 $f: \mathbb{R}^n \to \mathbb{R}^m$ 具有形式

$$f(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}, \qquad \boldsymbol{A} \in \mathbb{R}^{m\times n},\ \boldsymbol{b} \in \mathbb{R}^m,$$

则称 f 为仿射函数（线性变换 $+$ 平移）。



仿射函数保凸（定理）

定理

- 假设集合 $S \subseteq \mathbb{R}^n$ 是凸集，$f : \mathbb{R}^n \to \mathbb{R}^m$ 是仿射函数。则 S 在 f 下的象 $f(S) = \{f(\boldsymbol{x}) \mid \boldsymbol{x} \in S\} \subset \mathbb{R}^m$ 是凸集。
- 假设集合 $S \subseteq \mathbb{R}^n$ 是凸集，$f : \mathbb{R}^k \to \mathbb{R}^n$ 是仿射函数，则 S 在 f 下的原象 $f^{-1}(S) = \{\boldsymbol{x} \mid f(\boldsymbol{x}) \in S\} \subset \mathbb{R}^k$ 也是凸集。

例子：笛卡尔积与闵可夫斯基和

定理

- 如果 S_1 和 S_2 是凸集，直积或Cartesian乘积（笛卡尔积）

$$S_1 \times S_2 = \{(\boldsymbol{x}_1, \boldsymbol{x}_2) \mid \boldsymbol{x}_1 \in S_1, \boldsymbol{x}_2 \in S_2\} \text{ 也是凸集}$$

- 如果 S_1 和 S_2 是凸集，则它们的闵可夫斯基 *(Minkowski)* 和

$$S_1 + S_2 = \{\boldsymbol{x} + \boldsymbol{y} \mid \boldsymbol{x} \in S_1, \boldsymbol{y} \in S_2\} \text{也是凸集}$$

闵可夫斯基和保凸证明：先由第一条知 $S_1 \times S_2$ 凸。再注意 $S_1 + S_2$ 是它在仿射函数 $f(\boldsymbol{x}_1, \boldsymbol{x}_2) = \boldsymbol{x}_1 + \boldsymbol{x}_2$ 下的像，由上一页定理可知像集凸。

应用：利用保凸运算识别凸集

- 多面体 $P = \{\boldsymbol{x} \mid \boldsymbol{A}\boldsymbol{x} \preceq \boldsymbol{b}, \boldsymbol{C}\boldsymbol{x} = \boldsymbol{d}\}$
  - 把约束改写成“非负 + 等式”：构造仿射函数
  $$f(\boldsymbol{x}) = (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x},\ \boldsymbol{d} - \boldsymbol{C}\boldsymbol{x}).$$
  - 原象表达：$P = f^{-1}(\mathbb{R}_+^m \times \{0\})$。
  - 因为 $\mathbb{R}_+^m$ 与 $\{0\}$ 都凸，乘积也凸，所以其在仿射函数下的原象 P 凸。
- 椭球：$\mathcal{E} = \left\{\boldsymbol{x} \mid (\boldsymbol{x} - \boldsymbol{x}_c)^\top \boldsymbol{P}^{-1} (\boldsymbol{x} - \boldsymbol{x}_c) \leqslant 1\right\}$，其中 $\boldsymbol{P} \in \mathbf{S}_{++}^n$。
  - **观点 1（象）**：单位球 $B = \{\boldsymbol{u} \mid \|\boldsymbol{u}\|_2 \leqslant 1\}$ 在 $f(\boldsymbol{u}) = \boldsymbol{P}^{1/2}\boldsymbol{u} + \boldsymbol{x}_c$ 下的象 $\mathcal{E} = f(B)$。
  - **观点 2（原象）**：单位球 B 在 $g(\boldsymbol{x}) = \boldsymbol{P}^{-1/2} (\boldsymbol{x} - \boldsymbol{x}_c)$ 下的原象 $\mathcal{E} = g^{-1}(B)$。

结论：我们可以通过寻找这种映射关系来证明集合的凸性。

目录

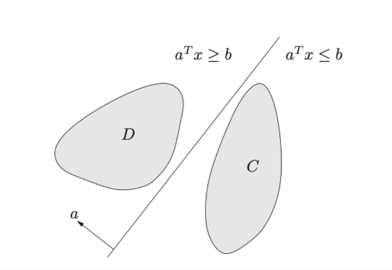
超平面分离定理

定理 (超平面分离定理)

假设 C 和 D 是两个不相交的凸集（即 $C \cap D = \emptyset$），那么存在 $\boldsymbol{a} \neq 0$ 和 $b \in \mathbb{R}$ 使得

$$\boldsymbol{a}^\top \boldsymbol{x} \leqslant b,\ \forall \boldsymbol{x} \in C; \qquad \boldsymbol{a}^\top \boldsymbol{x} \geqslant b,\ \forall \boldsymbol{x} \in D.$$

超平面 $\{\boldsymbol{x} \mid \boldsymbol{a}^\top \boldsymbol{x} = b\}$ 称为集合 C 和 D 的**分离超平面**。



为了把核心几何想法说清楚，下面先假设 $\mathrm{dist}(C, D) > 0$（此时可以严格分离）

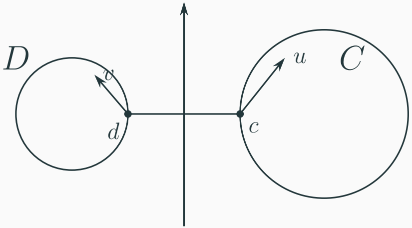
证明思路：寻找最近点（详见末尾选读）

假设 $\mathrm{dist}(C, D) > 0$。取距离最近的一对点 $\boldsymbol{c} \in C$ 和 $\boldsymbol{d} \in D$，使得

$$\|\boldsymbol{c} - \boldsymbol{d}\| = \min_{\boldsymbol{u} \in C, \boldsymbol{v} \in D} \|\boldsymbol{u} - \boldsymbol{v}\|$$

第 1 步：构造候选超平面（中垂面）：
超平面为 $\{\boldsymbol{x} \mid \boldsymbol{a}^\top \boldsymbol{x} = b\}$，其中

$$\boldsymbol{a} = \boldsymbol{d} - \boldsymbol{c}, \qquad b = \boldsymbol{a}^\top \frac{\boldsymbol{c} + \boldsymbol{d}}{2} = \frac{\|\boldsymbol{d}\|^2 - \|\boldsymbol{c}\|^2}{2}.$$



第 2 步：用“最近点”推出点积不等式： $\boldsymbol{a}^\top(\boldsymbol{v} - \boldsymbol{d}) \geq 0,\ \forall \boldsymbol{v} \in D$；$\boldsymbol{a}^\top(\boldsymbol{u} - \boldsymbol{c}) \leq 0,\ \forall \boldsymbol{u} \in C$。由此可得：$\forall \boldsymbol{v} \in D$ 有 $\boldsymbol{a}^\top \boldsymbol{v} - b \geq \frac{1}{2}\|\boldsymbol{d} - \boldsymbol{c}\|^2 > 0$，而 $\forall \boldsymbol{u} \in C$ 有 $\boldsymbol{a}^\top \boldsymbol{u} - b \leq -\frac{1}{2}\|\boldsymbol{d} - \boldsymbol{c}\|^2 < 0$，从而实现严格分离。

提示：

- 由于我们假设了 $\|\boldsymbol{d}-\boldsymbol{c}\|>0$，事实上对于任意 $\boldsymbol{u}\in C$, $\boldsymbol{a}^\top\boldsymbol{u}-b<0$；对于任意的 $\boldsymbol{v}\in D$, $\boldsymbol{a}^\top\boldsymbol{v}-b>0$，即该情况具有**严格分离**。
- 一般而言，不相交的凸集未必可以被严格分离（见作业）。

应用 1：闭凸集 = 半空间交

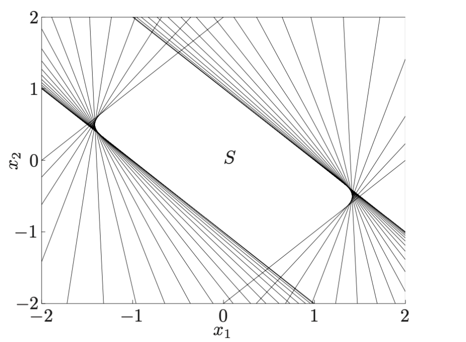
回顾半空间：$S(\boldsymbol{a}, b) = \{\boldsymbol{x} \mid \boldsymbol{a}^\top \boldsymbol{x} \le b\}$

定理

假设 C 是一个闭凸集，则 C 是所有包含它的半空间的交集。

$$C = \cap_{\boldsymbol{a}, b \in \Lambda} S(\boldsymbol{a}, b)$$

$$\Lambda = \{ (\boldsymbol{a}, b) \mid C \subseteq S(\boldsymbol{a}, b)\}$$



应用 2：超平面分离定理的逆定理

定理 (超平面分离定理的逆定理)

任何两个凸集 C 和 D，如果其中至少有一个是开集，那么

$$C \cap D = \emptyset \quad \Longleftrightarrow \quad \text{存在分离超平面。}$$

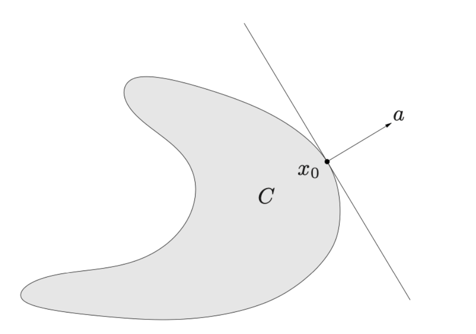
支撑超平面

定义 (支撑超平面)

对于集合 C 的边界上的点 $\boldsymbol{x}_0 \in \mathbb{R}^n$，若找到 $\boldsymbol{a} \in \mathbb{R}^n$ 使得

$$\boldsymbol{a}^\top \boldsymbol{x} \le \boldsymbol{a}^\top \boldsymbol{x}_0, \qquad \forall \boldsymbol{x} \in C$$

则超平面 $\{\boldsymbol{x} \mid \boldsymbol{a}^\top \boldsymbol{x} = \boldsymbol{a}^\top \boldsymbol{x}_0\}$ 被称为集合 C 在点 $\boldsymbol{x}_0$ 处的支撑超平面。



应用 3：支撑超平面定理

定理 (支撑超平面定理)

对于任意的非空凸集 C 和任意的边界点 $\boldsymbol{x}_0$，在 $\boldsymbol{x}_0$ 存在支撑超平面。

直观上：在凸集的边界“总能放一把尺子贴住它”，并且整个集合都在尺子的一侧。

目录

总结

- 理解并会用：交集运算、仿射映射是保凸运算
- 遇到新集合时：尝试用“保凸运算 + 已知凸集”来识别；或者回到定义验证
- 掌握超平面分离定理及其逆定理
- 知道支撑超平面定理的结论

阅读作业 & 参考资料：

- 课本第 2.3，2.5 章

超平面分离定理的详细补充证明（选读）

设 $\boldsymbol{a} = \boldsymbol{d} - \boldsymbol{c}$，$b = \boldsymbol{a}^\top \frac{\boldsymbol{c}+\boldsymbol{d}}{2}$，并定义 $f(\boldsymbol{x}) = \boldsymbol{a}^\top \boldsymbol{x} - b$。

(a) 对任意 $\boldsymbol{v} \in D$：我们考察 $f(\boldsymbol{v})$：

$$f(\boldsymbol{v}) = (\boldsymbol{d}-\boldsymbol{c})^\top \left(\boldsymbol{v} - \tfrac{\boldsymbol{c}+\boldsymbol{d}}{2}\right) = (\boldsymbol{d}-\boldsymbol{c})^\top(\boldsymbol{v}-\boldsymbol{d}) + \frac{\|\boldsymbol{d}-\boldsymbol{c}\|^2}{2} \geq \frac{\|\boldsymbol{d}-\boldsymbol{c}\|^2}{2} > 0.$$

其中关键是不等式 $(\boldsymbol{d}-\boldsymbol{c})^\top(\boldsymbol{v}-\boldsymbol{d}) \geq 0$。它来自“$\boldsymbol{d}$ 是 D 中到 $\boldsymbol{c}$ 的最近点”这一事实：

- 若点积 < 0（夹角为锐角），则沿着线段 $\boldsymbol{d} \to \boldsymbol{v}$ 走一小步，会让到 $\boldsymbol{c}$ 的距离变小。
- 这与“$\boldsymbol{d}$ 已经是最近点”矛盾。
- 因此点积只能 ≥ 0（夹角为直角或钝角）。

更确切地，对于任意的 $\boldsymbol{v} \in D$，我们可以定义函数

$$g(\theta) = \|\theta \boldsymbol{v} + (1-\theta)\boldsymbol{d} - \boldsymbol{c}\|^2, \qquad \theta \in [0,1].$$

由于 $\boldsymbol{d}$ 是 D 中到 $\boldsymbol{c}$ 的最近点，故 $g(\theta) \geq g(0) = \|\boldsymbol{d} - \boldsymbol{c}\|^2$，从而 $g'(0) \geq 0$。直接求导得

$$g'(0) = 2(\boldsymbol{d} - \boldsymbol{c})^\top (\boldsymbol{v} - \boldsymbol{d}),$$

于是 $(\boldsymbol{d} - \boldsymbol{c})^\top (\boldsymbol{v} - \boldsymbol{d}) \geq 0$，代回即可得到上式。

(b) 对任意 $\boldsymbol{u} \in C$:

$$f(\boldsymbol{u}) = (\boldsymbol{d}-\boldsymbol{c})^\top(\boldsymbol{u}-\boldsymbol{c}) - \frac{\|\boldsymbol{d}-\boldsymbol{c}\|^2}{2} \le -\frac{\|\boldsymbol{d}-\boldsymbol{c}\|^2}{2}$$

其中不等式类似可证明：令

$$g(\theta) = \|\theta\boldsymbol{u} + (1-\theta)\boldsymbol{c} - \boldsymbol{d}\|^2, \qquad \theta \in [0,1].$$

由于 $\boldsymbol{c}$ 是 C 中到 $\boldsymbol{d}$ 的最近点，故 $g(\theta) \ge g(0) = \|\boldsymbol{c}-\boldsymbol{d}\|^2$，从而 $g'(0) \ge 0$。直接求导得

$$g'(0) = 2(\boldsymbol{c}-\boldsymbol{d})^\top(\boldsymbol{u}-\boldsymbol{c}),$$

于是 $(\boldsymbol{d}-\boldsymbol{c})^\top(\boldsymbol{u}-\boldsymbol{c}) \le 0$，代回即可得到上式。