

第 2 章：数学基础

(1): 凸集

授课教师：曹语

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

背景

仿射/凸集/锥

重要例子

总结

背景和目标

有一类优化问题相对而言更容易求解：

$$\min_{x \in \Omega} f_0(x)$$

若 Ω 为凸集， f_0 为凸函数，则局部最优解即为全局最优解，且优化算法的效率能够得到保证。

LAGRANGE MULTIPLIERS AND OPTIMALITY

185

a convex set every locally optimal solution is global. Also, first-order necessary conditions for optimality turn out to be sufficient. A variety of other properties conducive to computation and interpretation of solutions ride on convexity as well. In fact the great watershed in optimization isn't between linearity and nonlinearity, but convexity and nonconvexity. Even for problems that aren't themselves of convex type, convexity may enter, for instance, in setting up subproblems as part of an iterative numerical scheme.

By the *convex case of problem (P)*, we'll mean the case where X is a convex set and, relative to X , the objective function f_0 and the inequality constraint functions f_1, \dots, f_s

事实上，优化问题的分水岭不是线性和非线性，而是凸性和非凸性。

R.T. Rockafellar, *SIAM Review*, 1993

本章主线： 在本章，我们依次回答下面三个问题：

- **什么是凸集？**（本节课先解决）
- 什么是凸函数？
- 如果 Ω 为凸集， f_0 为凸函数，最优解有什么性质？

提醒： 这节课的重点是先把“集合的形状”看清楚。

目录

背景

仿射/凸集/锥

重要例子

总结

直线、线段和射线

倘若 $\mathbf{x}_1 \neq \mathbf{x}_2$ 为 \mathbb{R}^n 空间中的两个点，先回顾三个最基本对象：

两点构成的直线（无限延伸）：

$$\mathbf{y} = \theta \mathbf{x}_1 + (1 - \theta) \mathbf{x}_2, \quad \theta \in \mathbb{R}$$

两点构成的线段（两点之间）：

$$\mathbf{y} = \theta \mathbf{x}_1 + (1 - \theta) \mathbf{x}_2, \quad \theta \in [0, 1]$$

通过某个点的射线（单向延伸）：

$$\mathbf{y} = \theta \mathbf{x}, \quad \theta \geq 0$$

定义 (仿射/Affine 的等价定义)

集合 C 是仿射的，等价地说，可以用下面几种方式来描述：

(1) (几何定义) 如果通过集合 $C \subseteq \mathbb{R}^n$ 中任意两个不同点的整条直线仍然都在集合 C 中。

(2) (代数定义) 若对任意的 $\mathbf{x}_1, \mathbf{x}_2 \in C$, $\theta \in \mathbb{R}$, 皆有 $\theta\mathbf{x}_1 + (1 - \theta)\mathbf{x}_2 \in C$ 。

(3) (推广) 若 $\mathbf{x}_1, \dots, \mathbf{x}_k \in C$, 且 $\theta_1 + \dots + \theta_k = 1$, 则

$$\sum_{i=1}^k \theta_i \mathbf{x}_i = \theta_1 \mathbf{x}_1 + \dots + \theta_k \mathbf{x}_k \in C$$

(4) (结构理解) $C = V + \mathbf{x}_0 = \{\mathbf{v} + \mathbf{x}_0 \mid \mathbf{v} \in V\}$, 其中 \mathbf{x}_0 是 C 中任意一点, V 是一个子空间 (即过原点的线性空间)。

仿射维数

直观理解：仿射集合就是平移后的线性子空间，典型例子是“不一定经过原点”的直线、平面。

基于该联系，可定义仿射集合 C 的维数为对应子空间 V 的维数。

例子：线性方程组的解 $\{x \mid Ax = b\}$ 是仿射的。

即线性等式约束所涉及到的集合

直观理解：仿射维数反映的是这个集合“本质上像几维的平面”。

仿射包

定义 由集合 $C \subseteq \mathbb{R}^n$ 中各点的所有仿射组合组成的集合，称为 C 的仿射包，记为 $\text{aff } C$ ：

$$\text{aff } C = \{\theta_1 \mathbf{x}_1 + \cdots + \theta_k \mathbf{x}_k \mid \mathbf{x}_1, \cdots, \mathbf{x}_k \in C, \theta_1 + \cdots + \theta_k = 1\}.$$

性质 $\text{aff } C$ 是包含 C 的最小仿射集合。例如：三角形的仿射包是它所在的整个平面。

仿射维数 集合 C 的仿射维数定义为 $\text{aff } C$ 的维数。

例子：在 \mathbb{R}^2 中，圆盘的仿射包是 \mathbb{R}^2 ，所以仿射维数为 2。

直观理解： $\text{aff } C$ 就是“把 C 放进一个尽量小的仿射集合里”。

定义 (凸集/convex set)

C 是凸集，核心判断标准是“任意两点连成的线段不跑出集合”。等价地：

(1) (几何直观) C 中任意两点间的线段仍然在 C 中。

(2) (代数定义) 对于任意 $\mathbf{x}_1, \mathbf{x}_2 \in C$ 和 $0 \leq \theta \leq 1$ 都有

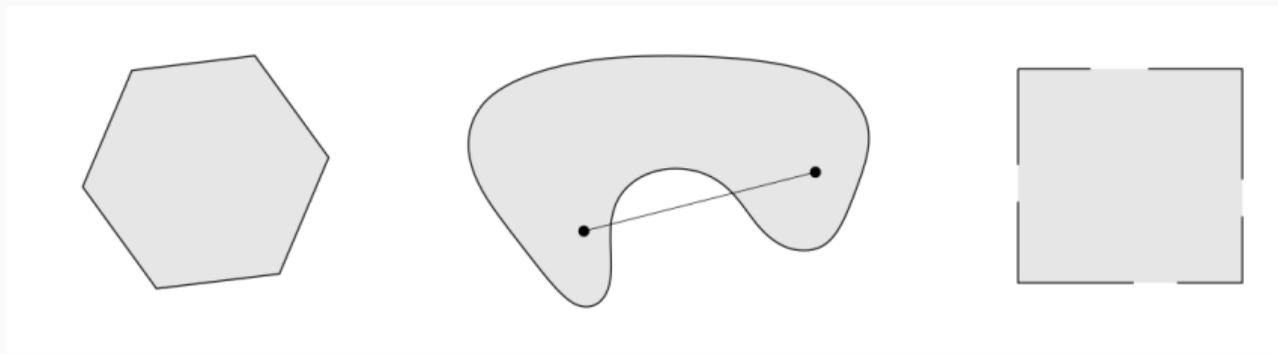
$$\theta \mathbf{x}_1 + (1 - \theta) \mathbf{x}_2 \in C$$

(3) (推广) 若 $\theta_i \geq 0$ 且 $\sum_{i=1}^k \theta_i = 1$ ，则 $\theta_1 \mathbf{x}_1 + \cdots + \theta_k \mathbf{x}_k \in C$ 。

【该线性组合 $\sum_{i=1}^k \theta_i \mathbf{x}_i$ 被称为 $\mathbf{x}_1, \cdots, \mathbf{x}_k$ 的一个凸组合】

判断凸集

例子：请判断如下的集合是否为凸集：



判断技巧： 找到两点，使它们连线有一部分落到集合外，即可判定非凸。

凸包

定义 (凸包/convex hull)

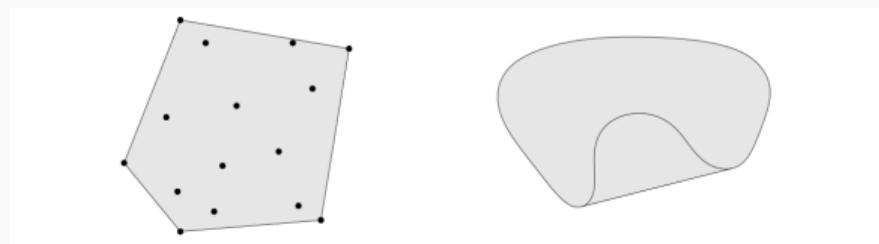
集合 C 中所有点的凸组合的集合被称为其**凸包**, 记为 $\text{conv } C$

$$\text{conv } C = \left\{ \theta_1 \mathbf{x}_1 + \cdots + \theta_k \mathbf{x}_k \mid \mathbf{x}_i \in C, \theta_i \geq 0, i = 1, \cdots, k, \sum_{i=1}^k \theta_i = 1 \right\}.$$

性质: C 的凸包是包含 C 的最小的凸集。

直观理解: 想象一个橡皮筋 (或保鲜膜) 紧紧包住集合 C 所形成的区域。

凸包的例子:

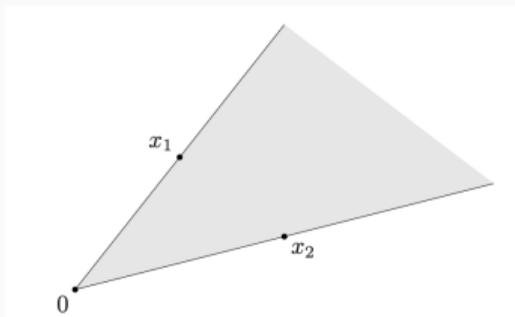


定义 (锥/Cone)

(1) 若对于任意 $x \in C$ 和 $\theta \geq 0$ 都有 $\theta x \in C$, 则称集合 C 是**锥**。

(即: 如果点 x 在集合中, 那么从原点出发经过 x 的整条射线都在集合中)

(2) 若集合 C 是锥, 且是凸集, 则称 C 为**凸锥**: 即对于任意 $x_1, x_2 \in C$ 和 $\theta_1, \theta_2 \geq 0$, 都有 $\theta_1 x_1 + \theta_2 x_2 \in C$



提示: 锥一定包含原点 (除非是空集)。

问题: 能否举出一个锥, 但它不是凸集?

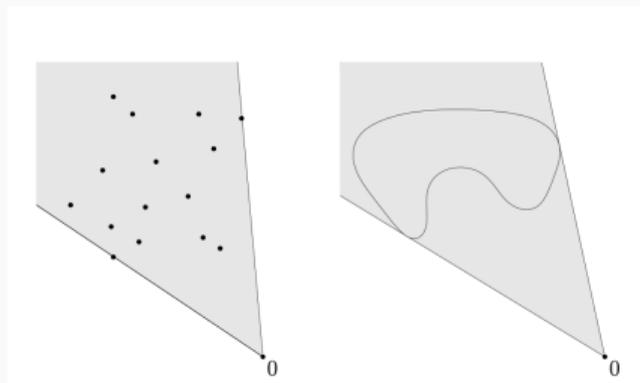
锥包

定义 (锥包)

集合 C 的锥包是 C 中元素的所有锥组合的集合，即

$$\{\theta_1 \mathbf{x}_1 + \cdots + \theta_k \mathbf{x}_k \mid \mathbf{x}_i \in C, \theta_i \geq 0, i = 1, \cdots, k\}$$

它是包含 C 的最小的凸锥。



小结：三个最基本对象

- 两点构成的直线：

$$\theta \mathbf{x}_1 + (1 - \theta) \mathbf{x}_2, \quad \theta \in \mathbb{R}$$

基于此，我们定义出仿射和仿射包

- 两点构成的线段：

$$\theta \mathbf{x}_1 + (1 - \theta) \mathbf{x}_2, \quad \theta \in [0, 1]$$

基于此，我们定义出凸集和凸包

- 通过某个点的射线：

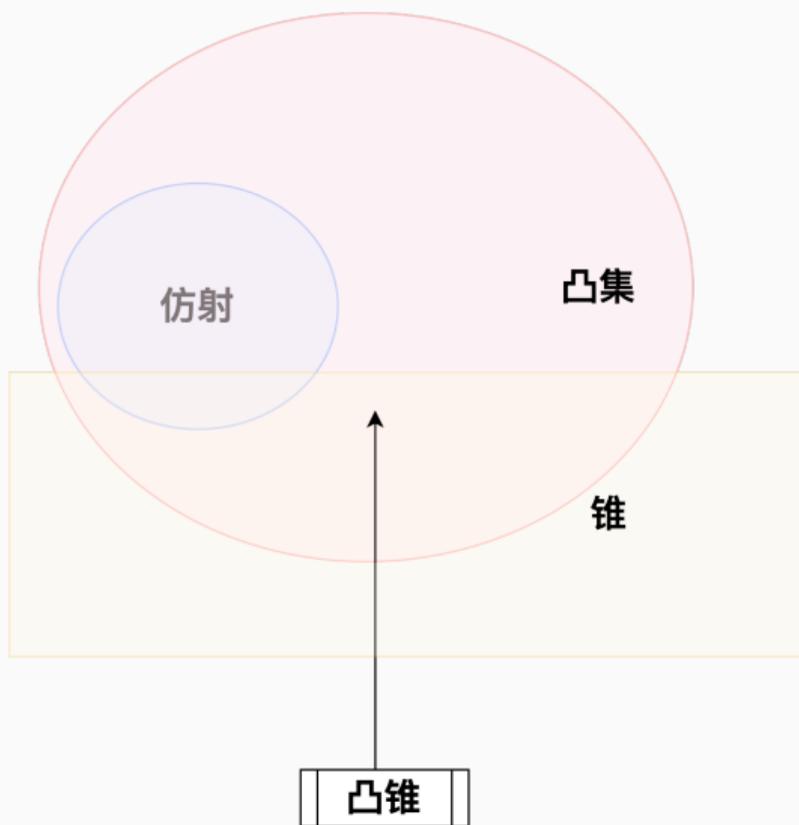
$$\theta \mathbf{x}, \quad \theta \geq 0$$

基于此，我们定义出锥和锥包

最值得记住的对应关系：直线 \leftrightarrow 仿射，线段 \leftrightarrow 凸集，射线 \leftrightarrow 锥。

概念之间的关系

多个概念之间的关联



练习

请判断如下的集合是否属于对应的 4 个类别：

集合	仿射	凸集	锥	凸锥
空集 \emptyset				
单点集 $\{x_0\}$				
不穿过原点的直线				
穿过原点的直线				

背景

仿射/凸集/锥

重要例子

总结

(1) 超平面和半空间

超平面：对于 $a \in \mathbb{R}^n$, $a \neq 0$, $b \in \mathbb{R}$,

$$\{x \mid a^\top x = b\}$$

分类：是仿射、是凸集、一般而言不是锥（除非 $b = 0$ ）。

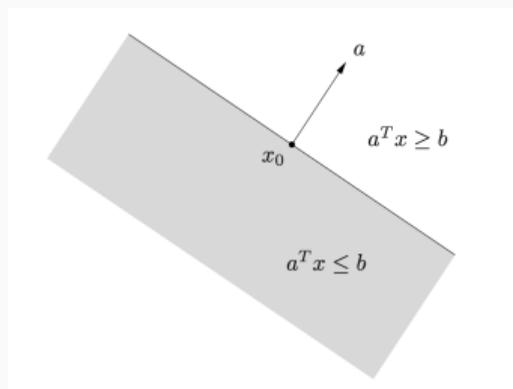
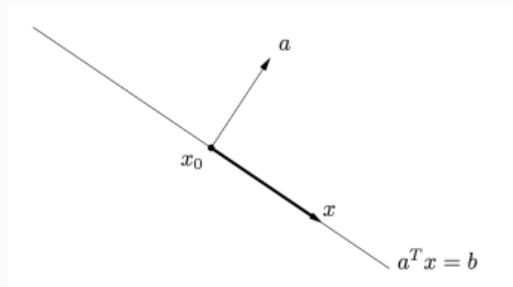
直观：它是空间中的“一张平的切面”。

(闭的) 半空间：对于 $a \in \mathbb{R}^n$, $a \neq 0$, $b \in \mathbb{R}$,

$$\{x \mid a^\top x \leq b\}$$

分类：不是仿射、是凸集、一般而言不是锥（除非 $b = 0$ ）。

直观：切开后的其中“一半”。



(2) Euclid 球和椭球

Euclid 球

$$\begin{aligned} B(\mathbf{x}_c, r) &= \{\mathbf{x} \mid \|\mathbf{x} - \mathbf{x}_c\|_2 \leq r\} = \left\{ \mathbf{x} \mid (\mathbf{x} - \mathbf{x}_c)^\top (\mathbf{x} - \mathbf{x}_c) \leq r^2 \right\} \\ &= \{\mathbf{x}_c + r\mathbf{u} \mid \|\mathbf{u}\|_2 \leq 1\} \end{aligned}$$

(直观理解：它是“以 \mathbf{x}_c 为中心、半径为 r 的球”。它是凸集，可由范数的三角不等式证明。)

椭球

$$\begin{aligned} \mathcal{E} &= \left\{ \mathbf{x} \mid (\mathbf{x} - \mathbf{x}_c)^\top \mathbf{P}^{-1} (\mathbf{x} - \mathbf{x}_c) \leq 1 \right\} \\ &= \left\{ \mathbf{x}_c + \mathbf{P}^{1/2}\mathbf{u} \mid \|\mathbf{u}\|_2 \leq 1 \right\} \end{aligned}$$

其中 $\mathbf{P} = \mathbf{P}^\top \succ 0$ ，即 \mathbf{P} 是对称正定矩阵。

(直观理解：椭球可以看成单位球经过平移、拉伸和旋转后得到)

- 更一般地可以定义**范数球**:

$$B(\mathbf{x}_c, r) = \{\mathbf{x} \mid \|\mathbf{x} - \mathbf{x}_c\| \leq r\}$$

注: 在 \mathbb{R}^2 中,

- 1-范数球是边长为 $\sqrt{2}$ 的正方形,
- ∞ -范数球是边长为 2 的正方形。

- 范数锥**:

$$C = \{(\mathbf{x}, t) \mid \|\mathbf{x}\| \leq t\} \subseteq \mathbb{R}^{n+1}$$

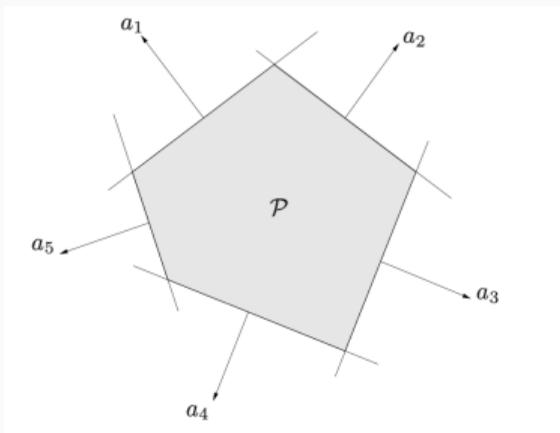
- 练习**: 证明范数锥是凸锥。

(3) 多面体/polyhedron

多面体是有限个半空间和超平面的交集，也就是一组线性不等式与线性等式的解集：

$$\mathcal{P} = \left\{ \mathbf{x} \mid \begin{array}{l} \mathbf{a}_j^\top \mathbf{x} \leq b_j, \quad j = 1, \dots, m, \\ \mathbf{c}_j^\top \mathbf{x} = d_j, \quad j = 1, \dots, p \end{array} \right\} \stackrel{\text{简写为}}{=} \left\{ \mathbf{x} \mid \mathbf{A}\mathbf{x} \preceq \mathbf{b}, \mathbf{C}\mathbf{x} = \mathbf{d} \right\}$$

即一般线性规划问题中最常见的可行域



$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1^\top \\ \vdots \\ \mathbf{a}_m^\top \end{bmatrix}, \quad \mathbf{C} = \begin{bmatrix} \mathbf{c}_1^\top \\ \vdots \\ \mathbf{c}_p^\top \end{bmatrix}$$

多面体为何是凸的？

由定义可知，超平面、半空间都是特殊的多面体。

练习： 证明多面体是凸集。(答案见课内)

(4) 对称矩阵集合

我们将 n 阶矩阵视作一个向量空间。如下是几个重要的凸集：

- n 阶对称矩阵的集合（是一个子空间）：

$$\mathbf{S}^n = \{ \mathbf{X} \in \mathbb{R}^{n \times n} \mid \mathbf{X} = \mathbf{X}^\top \}$$

- n 阶对称半正定矩阵（是凸锥）：

$$\mathbf{S}_+^n = \{ \mathbf{X} \in \mathbb{R}^{n \times n} \mid \mathbf{X} = \mathbf{X}^\top, \mathbf{X} \succeq 0 \}$$

（注意：这里的 $\succeq 0$ 表示特征值非负，而非每个元素非负）

- n 阶对称正定矩阵（是凸集，不是锥）：

$$\mathbf{S}_{++}^n = \{ \mathbf{X} \in \mathbb{R}^{n \times n} \mid \mathbf{X} = \mathbf{X}^\top, \mathbf{X} \succ 0 \}$$

目录

背景

仿射/凸集/锥

重要例子

总结

- 知道仿射、凸集、锥及其对应“包”的定义
- 掌握介绍过的 4 类重要例子，能判断它们分别属于哪些类别

阅读作业 & 参考资料:

- 课本第 2.1, 2.2 章

参考答案

- 第11页的例子：是，不是，不是
- 第13页的例子：平面直角坐标系的两个坐标轴的并集。
- 第17页的练习：

集合	仿射	凸集	锥	凸锥
空集 \emptyset	是	是	是	是
单点集 $\{x_0\}$	是	是	$x_0 = 0$ 时是 否则不是	$x_0 = 0$ 时是 否则不是
不穿过原点的直线	是	是	不是	不是
穿过原点的直线	是	是	是	是

参考答案 (续)

范数锥为何是凸锥：设 $C = \{(\mathbf{x}, t) \mid \|\mathbf{x}\| \leq t\}$. 要证明它是凸锥，只需分别验证“凸”与“锥”两件事：

- **凸性：** 任取 $(\mathbf{x}_1, t_1), (\mathbf{x}_2, t_2) \in C$ ，以及 $\theta \in [0, 1]$ ，由三角不等式可得

$$\|\theta\mathbf{x}_1 + (1 - \theta)\mathbf{x}_2\| \leq \theta\|\mathbf{x}_1\| + (1 - \theta)\|\mathbf{x}_2\| \leq \theta t_1 + (1 - \theta)t_2.$$

因此 $\theta(\mathbf{x}_1, t_1) + (1 - \theta)(\mathbf{x}_2, t_2) = (\theta\mathbf{x}_1 + (1 - \theta)\mathbf{x}_2, \theta t_1 + (1 - \theta)t_2) \in C$ 。

- **锥性：** 任取 $(\mathbf{x}, t) \in C$ 和 $\alpha \geq 0$ ，由范数的齐次性可得

$$\|\alpha\mathbf{x}\| = \alpha\|\mathbf{x}\| \leq \alpha t,$$

因此 $\alpha(\mathbf{x}, t) = (\alpha\mathbf{x}, \alpha t) \in C$ 。

综上， C 既是凸集，又满足锥性质，所以 C 是凸锥。