第九章 优化基础

第 29 讲 凸集

黄定江

DaSE @ ECNU
djhuang@dase.ecnu.edu.cn

29.1.1 直线与线段

- 设 $x_1 \neq x_2$ 为 R^n 空间中的两个点，那么具体有下列形式的点

$$y = \theta x_1 + (1-\theta)x_2, \theta \in \mathbb{R}$$

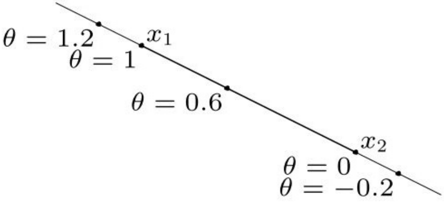
组成一条穿越 x_1 和 x_2 的直线。参数 $\theta = 0$ 对应 $y = x_2$，而 $\theta = 1$ 对应 $y = x_1$。参数 θ 的值在 0 和 1 之间变动，构成了 x_1 和 x_2 之间的 (闭) 线段。

直线与线段

- y 的表示形式

$$y = x_2 + \theta(x_1 - x_2)$$

给出了另一种解释：y 是基点 x_2(对应 $\theta = 0$) 和方向 $x_1 - x_2$(由 x_2 指向 x_1) 乘以参数 θ 的和。因此，θ 给出了由 x_2 通向 x_1 的路上的位置。当 θ 由 0 增加到 1，点 y 相应地由 x_2 移动到 x_1。如果 $\theta > 1$，点 y 在超越了 x_1 的直线上



29.1.2 仿射集合

- 如果通过集合 $\mathbb{C} \subseteq \mathbb{R}^n$ 中任意两个不同点的直线仍然在集合 $\mathbb{C}$ 是仿射的。也就是说，$\mathbb{C} \subseteq \mathbb{R}^n$ 是仿射的等价于：对于任意 $x_1, x_2 \in \mathbb{C}$ 及 $\theta \in \mathbb{R}$ 有 $\theta x_1 + (1-\theta)x_2 \in \mathbb{C}$。
- 这个概念可以扩展到多个点的情况。如果 $\theta_1 + \theta_2 + \cdots + \theta_k = 1$，我们称具有 $\theta_1 x_1 + \cdots + \theta_k x_k$ 形式为 $x_1, \cdots, x_k$ 的仿射组合。利用仿射集合的定义 (即仿射集合包含其中任意两点的仿射组合)，我们可以归纳出以下结论：一个仿射集合包含其中任意点的仿射组合，即如果 $\mathbb{C}$ 是一个仿射集合，$x_1, \cdots, x_k \in \mathbb{C}$, 并且 $\theta_1 + \cdots + \theta_k = 1$ ，那么 $\theta_1 x_1 + \cdots + \theta_k x_k$ 仍然在 $\mathbb{C}$ 中。

仿射集合

- 如果 $\mathbb{C}$ 是一个仿射集合并且 $x_0 \in \mathbb{C}$, 则集合

$$\mathbb{V} = \mathbb{C} - x_0 = \{x - x_0 | x \in \mathbb{C}\}$$

是一个子空间，即关于加法和数乘是封闭的。

- 因此，仿射集合 $\mathbb{C}$ 可以表示为

$$\mathbb{C} = \mathbb{V} + x_0 = \{v + x_0 | v \in \mathbb{V}\}$$

即一个子空间加上一个偏移。与仿射集合 $\mathbb{C}$ 相关联的子空间 $\mathbb{V}$ 与 x_0 的选取无关，所以 x_0 可以是 $\mathbb{C}$ 中的任意一点。我们定义仿射集合 $\mathbb{C}$ 的维数为子空间 $\mathbb{V} = \mathbb{C} - x_0$ 的维数，其中 x_0 是 $\mathbb{C}$ 中的任意元素。

仿射集合

- 我们称由集合 $\mathbb{C} \subseteq \mathbb{R}^n$ 中的点的所有仿射组合组成的集合为 $\mathbb{C}$ 的仿射包，记为 $\mathrm{aff}\mathbb{C}$：

$$\mathrm{aff}\mathbb{C} = \{\theta_1 x_1 + \cdots + \theta_k x_k | x_1, \cdots, x_k \in \mathbb{C}, \theta_1 + \cdots + \theta_k = 1\}$$

仿射包是包含 $\mathbb{C}$ 的最小的仿射集合，也就是说：如果 $\mathbb{S}$ 是满足 $\mathbb{C} \subseteq \mathbb{S}$ 的仿射集合，那么 $\mathrm{aff}\mathbb{C} \subseteq \mathbb{S}$

29.1.3 凸集

定义 1

- 集合 $\mathbb{C}$ 被称为凸集，如何 $\mathbb{C}$ 中任意两点间的线段仍在 $\mathbb{C}$ 中，即对于任意 $x_1, x_2 \in \mathbb{C}$ 和满足 $0 \leq \theta \leq 1$ 的 θ 都有

$$\theta x_1 + (1-\theta)x_2 \in \mathbb{C}$$

粗略地，如果集合中的每一点都可以被其他点沿着它们之间一条无阻碍的路径看见，那么这个集合就是凸集。所谓无阻碍，是指整条路径都在集合中。由于仿射集包含穿过集合中任意不同两点的整条直线，任意不同两点间的线段自然也在集合中。因而仿射集是凸集。

凸集

- 下图显示了 $\mathbb{R}^2$ 空间中一些简单的凸和非凸集合。

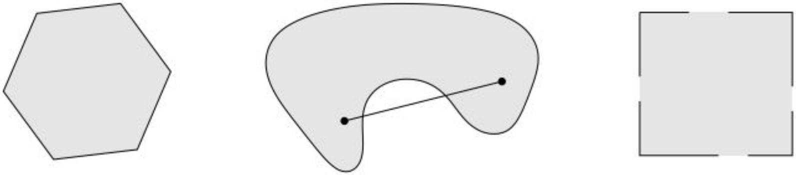


图 1: 左：包含边界的六边形是凸的。中：肾形集合不是凸的，因为图中所示集合中两点间的线段不为集合所包含。右：仅包含部分边界的正方形不是凸的。

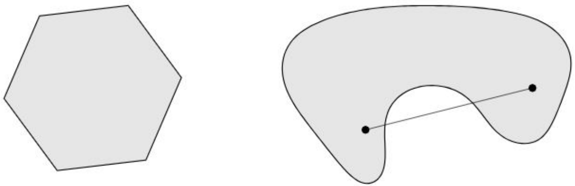
凸集

- 我们称集合 $\mathbb{C}$ 中所有点的凸组合的集合为其凸包，记为 conv$\mathbb{C}$，其中 $\theta_1 x_1 + \cdots + \theta_k x_k$ 为点 $x_1, \cdots, x_k$ 的一个凸组合，$\theta_1 + \cdots + \theta_k = 1$ 并且 $\theta_i \geq 0, i = 1, \cdots, k$:

$$\text{conv}\mathbb{C} = \{\theta_1 x_1 + \cdots + \theta_k x_k | x_i \in \mathbb{C}, \theta_i \geq 0, i = 1, \cdots, k, \theta_1 + \cdots + \theta_k = 1\}$$

  顾名思义，凸包总是凸的。它是包含 $\mathbb{C}$ 的最小的凸集。也就是说，如果 $\mathbb{B}$ 是包含 $\mathbb{C}$ 的凸集，那么 conv$\mathbb{C} \subseteq \mathbb{B}$。

下图显示了凸包的定义。



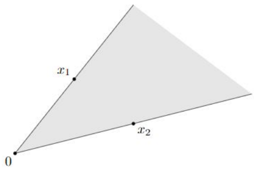
29.1.4 锥

定义 2

如果对于任意 $x \in \mathbb{C}$ 和 $\theta \geq 0$ 都有 $\theta x \in \mathbb{C}$, 我们称集合 $\mathbb{C}$ 是锥或者非负齐次。如果集合 $\mathbb{C}$ 是锥，并且是凸的，则称 $\mathbb{C}$ 是凸锥，即对于任意 $x_1, x_2 \in \mathbb{C}$ 和 $\theta_1, \theta_2 \geq 0$, 都有

$$\theta_1 x_1 + \theta_2 x_2 \in \mathbb{C}$$

在几何上，具有此类形式的点构成了二维的扇形，这个扇形以 0 为定点，边通过 x_1 和 x_2，如下图所示：



锥

- 具有 $\theta_1 x_1 + \cdots + \theta_k x_k, \theta_1, \cdots, \theta_k \geq 0$ 形式的点称为 $x_1, \cdots, x_k$ 的锥组合 (或非负线性组合)。如果 x_i 均属于凸锥 $\mathbb{C}$, 那么，x_i 的每一个锥组合也在 $\mathbb{C}$ 中。反言之，集合 $\mathbb{C}$ 是凸锥的充要条件是它包含其元素的所有锥组合。

29.1.5 凸集的重要例子

首先介绍一些简单的例子。

- 空集, 任意一个点 (即单点集)$\{x_0\}$ 、全空间 $\mathbb{R}^n$ 都是 $\mathbb{R}^n$ 的仿射 (自然也是凸的) 子集。
- 任意直线是仿射的。如果直线通过零点，则是子空间，因此，也是凸锥。
- 一条线段是凸的，但不是仿射的 (除非退化为一个点〉。
- 一条射线，即具有形式 $\{x_0 + \theta v | \theta \geq 0\}, v \neq 0$ 的集合，是凸的，但不是仿射的。如果射线的基点 x_0 是 0，则它是凸锥。
- 任意子空间是仿射的、凸锥 (自然是凸的)。

超平面与半空间

定义 3

超平面是具有下面形式的集合

$$\{x|a^Tx = b\}$$

其中 $a \in \mathbb{R}, a \neq 0$ 且 $b \in \mathbb{R}$。

- 解析地，超平面是关于 x 的非平凡线性方程的解空间 (因此是一个仿射集合)。
- 几何上，超平面 $\{x|a^Tx = b\}$ 可以解释为与给定向量 a 的内积为常数的点的集合; 也可以看成法线方向为 a 的超平面，而常数 $b \in \mathbb{R}$ 决定了这个平面从原点的偏移。

- 为更好地理解这个几何解释，可以将超平面表示成下面的形式
$$\{x|a^T(x-x_0)=0\}$$
其中 x_0 是超平面上任意一点（既满足 $a^Tx_0=b$ 的点）。进一步可表示为
$$\{x|a^T(x-x_0)=0\}=x_0+a^{\perp}$$
这里 a 表示 a 的正交补，即与 a 正交的向量的集合：
$$a=\{v|a^Tv=0\}$$
- 这些几何解释可见下图：

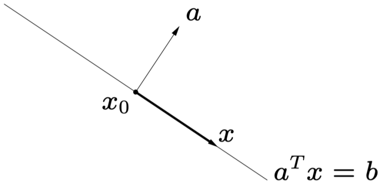


图 3: $\mathbb{R}^2$ 中由法向量 a 和超平面上一点 x_0 确定的超平面。对于超平面上任意一点 x，$x-x_0$（如深色箭头所示）都垂直于 a。

超平面与半空间

- 一个超平面将 $\mathbb{R}^n$ 划分为两个半空间，其中 $a \neq 0$。半空间是凸的，但不是仿射的，如下图所示：

超平面与半空间

- 半空间也可以表示为，

$$\{x|a^T(x-x_0)\le 0\}$$

- 其中 x_0 是相应超平面上的任意一点，即 x_0 满足 $a^Tx_0=b$。

$$\{x|a^T(x-x_0)\le 0\}$$

有一个简单的几何解释：半空间由 x_0 加上任意与（向外的法）向量 a 呈钝角（或直角）的向量组成。

- 半空间的边界是超平面 $\{x|a^Tx=b\}$ 。集合 $\{x|a^Tx<b\}$ 是半空间 $\{x|a^Tx\le b\}$ 的内部，称为开半空间。

Euclid 球

- $\mathbb{R}^n$ 中的空间 $Euclid$ 球 (或简称为球) 具有下面的形式，

$$B(x_c, r) = \{x| \|x - x_c\|_2 \leq r\} = \{x|(x - x_c^T)(x - x_c) \leq r^2\}$$

其中 $r \geq 0$,$\|\|_2$ 表示 $Euclid$ 范数，即 $\|u\|_2 = (u^T u)^{\frac{1}{2}}$。向量 x_c 是球心，标量 r 为半径。$B(x_c, r)$ 由距离球心 x_c 距离不超过 r 的所有点组成。$Euclid$ 球的另一个常见表达式为

$$B(x_c, r) = \{x_c + ru| \|u\|_2 \leq 1\}$$

Euclid 球

- *Euclid* **球是凸集，即如果** $\|x_1 - x_c\|_2 \le r, \|x_2 - x_c\|_2 \le r$，**并且** $0 \le \theta \le 1$, **那么**

$$\begin{aligned}\|\theta x_1 + (1-\theta)x_2 - x_c\|_2 &= \|\theta(x_1 - x_c) + (1-\theta)(x_2 - x_c)\|_2 \\ &\le \theta\|(x_1 - x_c)\|_2 + (1-\theta)\|(x_2 - x_c)\|_2 \\ &\le r\end{aligned}$$

范数球与范数锥

- 设 $\|.\|$ 是 $\mathbb{R}^n$ 中的范数。由范数的一般性质可知，以 r 为半径，x_c 为球心的范数球 $\{x|\|x-x_c\| \leq r\}$ 是凸的。
- 关于范数 $\|.\|$ 的范数锥是集合

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

顾名思义，它是一个凸锥。

多面体

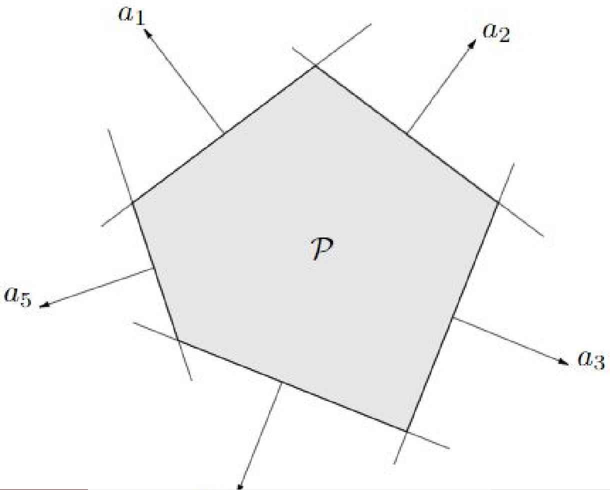
- 多面体被定义为有限个线性等式和不等式的解集。

$$P = \{x | a_j^T x \leq b_j, j = 1, \cdots, m, c_j^T x = d_j, j = 1, \cdots, p\}$$

因此，多面体是有限个半空间和超平面的交集。仿射集合（例如子空间、超平面、直线)、射线、线段和半空间都是多面体。显而易见，多面体是凸集。有界的多面体有时也称为多胞形，但也有一些作者反过来使用这两个概念 (即用多胞形表示具有上面形式的集合，而当其有界时称为多面体)。

多面体

下图显示了一个由五个半空间的交集定义的多面体。



半正定锥

- 我们用 S^n 表示对称 $n \times n$ 矩阵的集合，即

$$S^n = \{X \in R^{n\times n} | X = X^T\}$$

这是一个维数为 $n(n+1)/2$ 的向量空间。我们用 S_+^n 表示对称半正定矩阵的集合：

$$S_+^n = \{X \in S^n | X \succeq 0\}$$

用 S_{++}^n 表示对称正定矩阵集合：

$$S_{++}^n = \{X \in S_+^n | X \succ 0\}$$

半正定锥

- 集合 S_+^n 是一个凸锥：如果 $\theta_1, \theta_2 \geq 0$ 并且 $A, B \in S_+^n$ ，那么 $\theta_1 A + \theta_2 B \in S_+^n$。从半正定矩阵的定义可以直接得到：对于任意 $x \in Rn$ 如果 $A \geq 0, B \geq 0$，那么，我们有
$$x^T(\theta_1 A + \theta_2 B)x = \theta_1 x^T Ax + \theta_2 x^T Bx \geq 0$$

保凸运算

本节将描述一些保凸运算，利用它们，我们可以从凸集构造出其他凸集。

- 交集
- 仿射函数
- 线性分式及透视函数

交集

- 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. 这个性质可以扩展到无穷个集合的交: 如果对于任意 $\alpha \in \mathcal{A}$，S_α 都是凸的，那么 $\cap_{\alpha\in\mathcal{A}} S_\alpha$ 也是凸集。(子空间、仿射集合和凸锥对于任意交运算也是封闭的。)
- 作为一个简单的例子，多面体是半空间和超平面 (它们都是凸集) 的交集，因而是凸的。

交集

例 1

- 半正定锥 S_+^n 可以表示为，

$$\cap_{z\neq 0}\{X \in S^n | z^T Xz \geq 0\}.$$

对于任意 $z \neq 0$，$z^T Xz$ 是关于 X 的 (不恒等于零的) 线性函数，因此集合

$$\{X \in S^n | z^T Xz \geq 0\}$$

实际上就是 S^n 的半空间。由此可见，半正定锥是无穷个半空间的交集，因此是凸的.

交集

例 2

- 考虑集合
$$S=\{x\in\mathbb{R}^m|\,|p(t)|\le 1\text{对于}\,|t|\le\pi/3\}$$
其中 $p(t)=\sum_{k=1}^m x_k\cos kt$。集合 S 可以表示为无穷个平板的交集:$S=\cap_{|t|\le\pi/3}S_t$，其中
$$S_t=\{x|-1\le(\cos t,\cdots,\cos mt)^T x\le 1\}$$
因此，S 是凸的。

交集

- 对于 $m=2$ 的情况，它的定义和集合可见图4和图5。

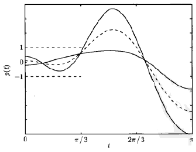


图 4: 对应于 $m=2$ 中的点的三角多项式. 虚线所示的三角多项式是另外两个的平均.

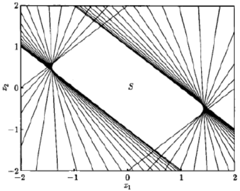


图 5: 图中央的白色区域显示了 m = 2 情况下例定义的集合 s. 这个集合是无限多个 (图中显示了其中 20 个) 平板的交集，所以是凸的。

仿射函数

- 函数 $f:\mathbb{R}^n\to\mathbb{R}^m$ 是仿射的，如果它是一个线性函数和一个常数的和，即具有 $f(x)=Ax+b$ 的形式，其中 $A\in\mathbb{R}^{m\times n}, b\in\mathbb{R}^m$。假设 $S\subseteq\mathbb{R}^n$ 是凸的，并且 $f:\mathbb{R}^n\to\mathbb{R}^m$ 是仿射函数。那么，S 在 f 下的象

$$f(S)=\{f(x)|x\in S\}$$

是凸的。类似的，如果 $f:\mathbb{R}^k\to\mathbb{R}^n$ 是仿射函数。那么 S 在 f 下的原象

$$f^{-1}(S)=\{x|f(x)\in S\}$$

是凸的。

仿射函数

- 两个简单的例子是伸缩和平移。如果 $S \subseteq \mathbb{R}^n$ 是凸集，$\alpha \in \mathbb{R}$ 并且 $a \in \mathbb{R}^n$，那么，集合 αS 和 $S + a$ 是凸的，其中

$$\alpha S = \{\alpha x | x \in S\}, \qquad S + a = \{x + a | x \in S\}$$

- 一个凸集向它的某几个坐标的投影是凸的，即：如果 $S \subseteq \mathbb{R}^m \times \mathbb{R}^n$ 是凸集，那么

$$T = \{x_1 \in \mathbb{R}^m | (x_1, x_2) \in S \text{对于某些} x_2 \in \mathbb{R}^n\}$$

是凸集。

仿射函数

- 两个集合的和可以定义为:

$$S_1 + S_2 = \{x \in S_1, y \in S_2\}$$

如果 S_1 和 S_2 是凸集，那么，$S_1 + S_2$ 是凸的。可以看出，如果 S_1 和 S_2 是凸的，那么其直积或 Cartesian 乘积

$$S_1 \times S_2 = \{(x_1, x_2) | x_1 \in S_1, x_2 \in S_2\}$$

也是凸集，这个集合在线性函数 $f(x_1, x_2) = x_1 + x_2$ 下的像是和 $S_1 + S_2$。

仿射函数

- 我们也可以考虑 $S_1, S_2 \in \mathbb{R}^n \times \mathbb{R}^m$ 的部分和，定义为

$$S = \{(x, y_1 + y_2) | (x, y_1) \in S_1, (x, y_2) \in S_2\}$$

其中 $x \in \mathbb{R}^n, y_i \in \ \mathbb{R}^m$。$m = 0$ 时，部分和给出了 S_1 和 S_2 的交集；$n = 0$，部分和等于集合之和。凸集的部分和是凸集。

线性分式及透视函数

定义 4

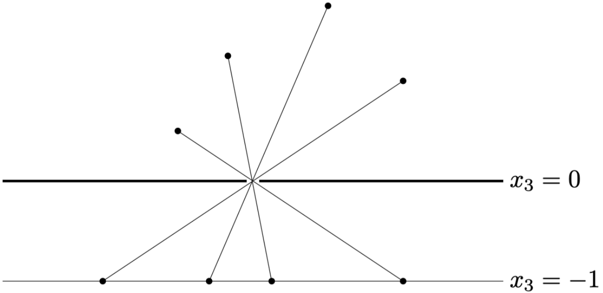
我们定义 $P:\mathbb{R}^{n+1}\to\mathbb{R}^n, P(z,t)=z/t$ 为透视函数，其定义域为 $\mathbf{dom}\,P=\mathbb{R}^n\times\mathbb{R}_+$。透视函数对向量进行伸缩，或称为规范化，使得最后一维分量为 1 并舍弃之。

- 如果 $C\subseteq\mathbf{dom}\,P$ 是凸集，那么它的象

$$P(C)=\{P(x)|x\in C\}$$

也是凸集。这个结论很直观：通过小孔观察一个凸的物体，可以得到凸的象。为解释这个事实，我们将说明在透视函数作用下，线段将被映射成线段。

- 我们用小孔成像来解释透视函数。($\mathbb{R}^3$ 中的)小孔照相机有一个不透明的水平面 $x_3=0$ 和一个在原点的小孔组成，光纤透过这个小孔在 $x_3=-1$ 呈现出一个水平图像。在相机上面 $x(x_3>0)$ 出的一个物体，在相平面的点 $-(x_1/x_3, x_2/x_3, 1)$ 处形成一个图像。忽略像点的最后一维分量（因为它恒等于 -1），x 处的点的像在像平面上呈现于 $y=-(x_1/x_3, x_2/x_3)=-P(x)$ 处。



线性分式及透视函数

- 一个凸集在透视函数下的原象也是凸的: 如果 $C \subseteq \mathbb{R}^n$ 为凸集，那么

$$P^{-1}(C) = \{(x, t) \in \mathbb{R}^{n+1} | x/t \in C, t > 0\}$$

是凸集。

线性分式函数

- 线性分式函数由透视函数和仿射函数复合而成。设 $g:\mathbb{R}^n \to \mathbb{R}^{m+1}$ 是仿射的，即

$$g(x) = \begin{bmatrix} A \\ c^T \end{bmatrix} x + \begin{bmatrix} b \\ d \end{bmatrix}$$

其中 $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m, c \in \mathbb{R}^n$ 并且 $d \in \mathbb{R}$。则由 $f = P \circ g$ 给出的函数 $f:\mathbb{R}^n \to \mathbb{R}^m$

$$f(x) = (Ax+b)/(c^Tx+d), \mathbf{dom}\, f = \{x | c^Tx+d>0\}$$

称为线性分式 (或投射) 函数。如果 $c=0, d>0$。则 f 的定义域为 $\mathbb{R}^n$，并且 f 是仿射函数。因此，我们可以将仿射和线性函数视为特殊的线性分式函数。

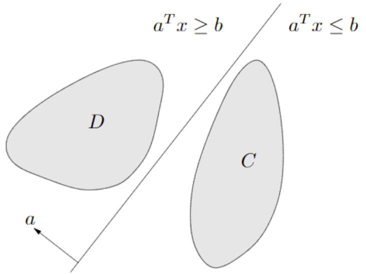
线性分式函数

- 类似于透视函数，线性分式函数也是保凸的。如果 C 是凸集并且在 f 的定义域中（即任意 $x \in C$ 满足 $c^T x + d > O$），那么 C 的象 $f(C)$ 也是凸集。根据前述的结果可以直接得到这个结论:C 在仿射映射下的象是凸的，并且在透视函数 P 下的映射 (即 $f(C)$) 是凸的。类似地，如果 $C \subseteq R^m$ 是凸集，那么其原象 $f^{-1}(C)$ 也是凸的。

分离与支撑超平面

定理 1

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 $a \neq 0$ 和 b 使得对于所有 $x \in C$ 有 $a^T x \leq b$, 对于所有 $x \in D$ 有 $a^T \geq b$。换言之，仿射函数 $a^T x - b$ 在 C 中非正，而在 D 中非负，超平面 $\{x | a^T x = b\}$ 称为集合 C 和 D 的分离超平面，或称超片面分离了集合 C 和 D。



支撑超平面

- 设 $C \subseteq R^n$ 而 x_0 是其边界 **bd**C 上的一点，即

$$x_0 \in \mathbf{bd}\, C = \mathbf{cl}\, C \backslash \mathbf{int}\, C.$$

如果 $a \neq 0$. 并且对任意 $x \in C$ 满足 $a^T x \leq a^T x_0$. 那么称超平面 $\{x | a^T x = a^T x_0\}$ 为集合 C 在点 X_0 处的支撑超平面。这等于说点 x_0 与集合 C 被超平面所分离。其几何解释是超平面 $\{x | a^T x = a^T x_0\}$ 与 C 相切于点 x_0，而且半空间 $\{x | a^T x \leq a^T x_0\}$ 包含 C。