# 通过支撑超平面定理的另一种证明实现非凸分离

Ali Taherinassaj* Yiling Chen*ali
taherinassaj@g.harvard.edu yiling@seas.harvard.edu

2023年10月1日

arXiv:2303.05423v2 [math.OC] 8 Oct 2023

**摘要**

我们给出了支撑超平面定理的一个简明证明。随后我们发现，该证明不仅证明了支撑超平面定理，还将其推广为适用于某些非凸集的超平面分离定理。其核心思想是：如果从某个点的视角看去，两个集合呈现为凸集，那么这两个集合便可由一个超平面分离。这种全局拓扑与局部向量空间之间的联系，为未来将结果推广至流形提供了启示。该结果或许也具有教学价值，因为它重新组织了凸优化和分析教材中通常呈现支撑超平面定理及其依赖关系的结构。

## 1 引言

机器学习、经济学和工程学中的许多问题通常都涉及优化算法的使用。凸分析具有特殊的理论意义，因为它为分析优化算法的收敛速度提供了一个框架，而分离超平面定理是这一框架的核心。此外，支持向量机等算法 [1], 会直接利用超平面对数据进行分类。

凸优化教材，例如 [4], ，通常遵循这样的顺序：先证明分离超平面定理，再以此为基础建立支撑超平面定理和 Farkas 引理。然而，一个值得注意的观察是，Farkas 引理是一个更简单的定理，可以通过 Fourier–Motzkin 消元法以线性代数的方法证明（见 [2]），而且它似乎先于适用于任意集合的分离定理。为了探寻更基础的理论依据，我们尝试通过一种不同且可能更直观的方法来证明该定理。

稍作重新组织后，便可将分离超平面定理归结为支撑超平面定理，正如 [3] 中所做的那样；他们通过研究序列并运用投影定理来证明支撑超平面定理。我们进一步将适用于任意集合的支撑超平面定理归结为适用于凸锥的支撑超平面定理。尽管这一步小小的改动在欧几里得空间中并无影响，却可能使我们能够通过点集的视锥[1] 来处理这些点集；这些视锥是流形 $M$上某一点 $p$ 的切空间 $T_pM$ 的子集。

## 2 符号与预备知识

全文中，我们用 int $(S)$ 表示集合 $S$ 的内部，用 cl $(S)$ 表示其闭包，用 bd $(S)$ 表示其边界。对于两个集合 $A$ 和 $B$，$A \setminus B$ 表示集合差，$A \pm B$ 表示闵可夫斯基加法／减法。现在我们定义几个将在主要结果中用到的工具。

**定义 2.1**（视锥）。设 $S \subseteq \mathbb{R}^n$ 是一个集合，$p \in \mathbb{R}^n$ 是一个点。那么，从 $p$ 的视角看 $S$ 的锥，定义为尖端位于 $p$ 且包含 $S$ 的最小锥，即

$$\operatorname{cone}_p(S) := \{v \in \mathbb{R}^n \mid \exists\, t_v \in \mathbb{R}_{\geq 0} : t_v v \in S_p\}. \tag{1}$$

*哈佛大学，美国马萨诸塞州剑桥[1]在最短测地线唯一的 Hadamard流形中，两者可通过对数映射相互转换。

其中 $S_p$ 表示 $S - p$。

**注 2.2.** 当点 $p$属于 $S$时，视锥变得平凡，即 .

$$\mathrm{cone}_p(S) = \mathbb{R}^n. \tag{2}$$

# 3 主要结果

**定理 3.1 (**支撑超平面定理**).** 设 $C \subseteq \mathbb{R}^n$ 是一个凸集， $p \notin \mathrm{int}(C)$是一个位于 $C$内部之外的点。存在一个经过 $p$ 的超平面，使得 $C$ 包含在其某个半空间内。

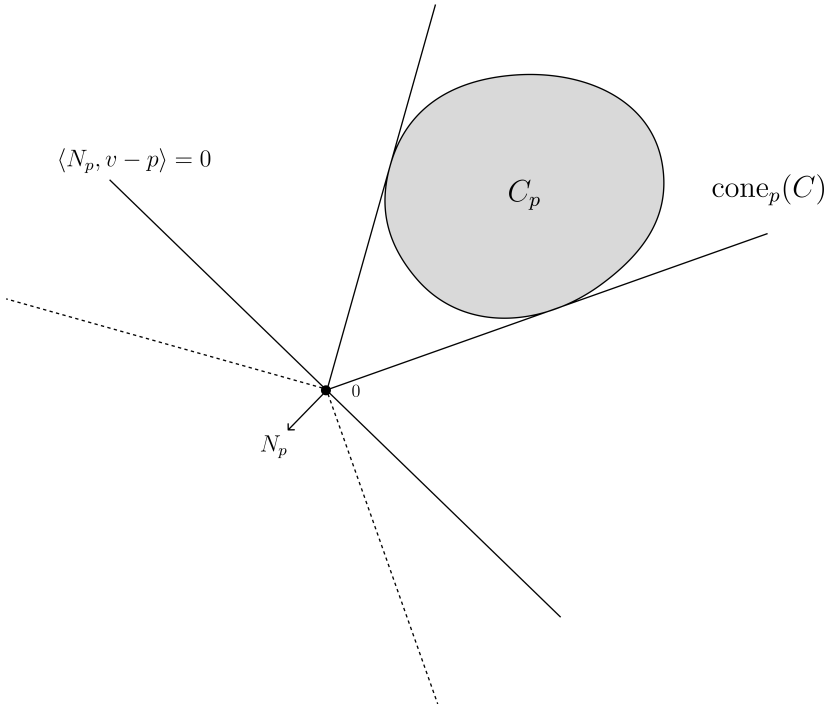


图 1：支撑超平面定理（定理 3.1）的示意图。

证明. 假设 $C$ 是 $\mathbb{R}^n$的一个真子集且非空[2] 。 $C_p$ 是锥$_p$($C$)的子集，而根据引理 5.4，锥$_p$($C$)本身是凸的；根据引理 5.6，由于锥$_p$($C$)具有凸性，它是某个半空间的子集。

■

**注 3.2.** 经过点 $p$ 的半空间可由法向量 $N$定义 $_p$:

$$\exists N_p \in \mathbb{R}^n \setminus \{0\} : \forall v \in C : \langle N_p, v - p\rangle \leq 0. \tag{3}$$

**注 3.3.** 请注意，在定理 *3.1* 的证明中，我们只使用了 $C$ 的凸性来推出锥$_p$($C$)的凸性。如果锥$_p$($C$) 是凸的，而 $C$ 本身不一定是凸的，那么定理 *3.1* 仍然成立。

**定理 3.4 (**广义分离超平面定理**).** 设 $A$ 和 $B$ 是 $\mathbb{R}^n$ 的两个任意子集。若存在一点 $p$ ，使得两个视锥锥$_p$($A$) 和锥$_p$($B$) 均为凸集且内部不相交，那么存在一个超平面将这两个集合分开。

证明。 我们可以像 [3] 中那样，从原点$p$:= 0 对 $C$ := 锥$_p$(*A*) –锥$_p$(*B*) 应用支撑超平面定理。集合 $C$ 是凸的，因为两个凸锥的闵可夫斯基和是凸的。

[2]如果 $C$ 是整个空间，那么一开始就不存在位于 $C$ 内部之外的点 $p$ 。同样，如果 $C$ 是空集，那么它总是任意半空间的子集。无论哪种情况，该定理都平凡成立。

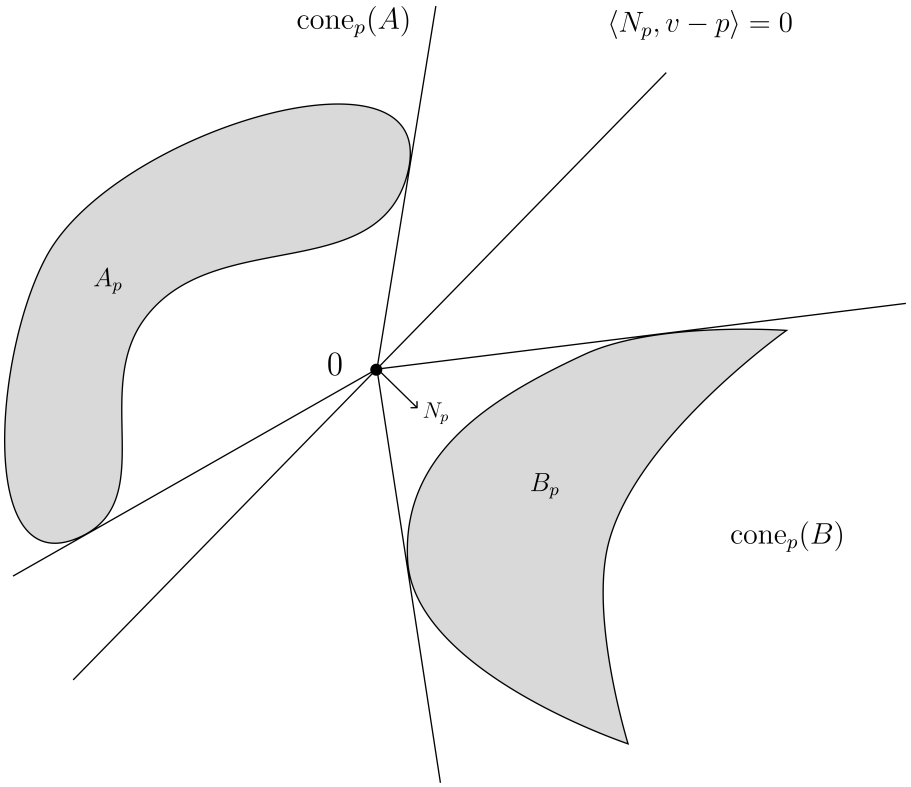


图 2：只要存在一个点 $p$，从该点的视角看，$A$ 和 $B$ 是凸集，就可能有多个超平面将 $A$ 和 $B$ 分隔开。

此外，原点位于锥$_p$($A$) – 锥$_p$($B$) 的边界上。因此存在一个超平面，即

$$\exists N_p \in \mathbb{R}^n \setminus \{0\} : \forall a \in A : \forall b \in B : \langle N_p, a\rangle \leq \langle N_p, b\rangle. \tag{4}$$

■

# 4 致谢

A.T. 感谢 Nima Fasih 提供有益的交流并帮助校对。此项工作是哈佛大学 CS 128：凸优化及其在机器学习中的应用课程开发的一部分，该课程由 Yiling Chen 博士讲授。

# 参考文献

[1] Corinna Cortes 和 Vladimir Vapnik。“支持向量网络”。载于：机器学习 20.3（1995），第 273–297 页。doi：10.1007/BF00994018。url：https://doi.org/10.1007/BF00994018。[2] G. M. Ziegler。多胞体讲义。第 1 版。数学研究生教材。施普林格纽约，1995，第 XI、370 页。isbn：978-0-387-94329-9。doi：10.1007/978-1-4613-8431-1。url：https: // doi.org/10.1007/978-1-4613-8431-1。[3] Dimitri P. Bertsekas、Angelia Nedi´c 和 Asuman E. Ozdaglar。凸分析与优化。2003。[4] Stephen Boyd 和 Lieven Vandenberghe。凸优化。剑桥大学出版社，2004。doi：10.1017/CBO9780511804441。[5] Mikhail Lavrov。第 *19* 讲：*Fourier–Motzkin* 消元法。Math 482：线性规划讲义。2019。url：https://faculty.math.illinois.edu/~mlavrov/docs/482-fall-2019/lecture19.pdf。

# 5 附录

这里回顾一下锥和凸性的基础知识。

**定义 5.1**（锥）。 当集合 $K \subseteq \mathbb{R}^n$ 满足以下条件时，称其为锥：

$$\forall v \in K : \forall \theta \in \mathbb{R}_{>0} : \theta v \in K \tag{5}$$

**引理 5.2**（凸锥）。 锥 $K \subseteq \mathbb{R}^n$ 是凸的，当且仅当：

$$\forall v, w \in K : v + w \in K \tag{6}$$

证明。 对于正向推导，若 $v$ 和 $w$ 属于该锥，则根据定义 5.1， $2v$ 和 $2w$ 也属于该锥。由凸性可知，两者的平均值，即 $v + w$，也必定属于该锥。

对于反向推导，我们必须证明：对于任意两个向量 $v$ 和 $w$，以及任意 $\theta \in [0, 1]$，在任意两个向量之和都属于该锥的前提下，加权平均 $\theta v + (1 - \theta)w$ 也属于该锥。 $\theta = 0$ 和 $\theta \in (0,1)$ 的情形必须分别处理。当 $\theta = 0$ 时，加权组合只有借助额外假设——我们起初选取的向量属于该锥——才会落在锥中。对于另一种情形，我们可以将 $v$ 按 $\theta > 0$ 缩放，并将 $w$ 按 $(1 - \theta)$ 缩放，同时仍使这两个向量属于该锥，从而将它们相加得到 $\theta v + (1 - \theta)w$。■

**定义 5.3**（锥包）。 集合 $S \subseteq \mathbb{R}^n$ 的锥包定义如下：

$$\text{conic-hull}\,(S) := \{\theta_v \cdot v + \theta_w \cdot w \,|\, v, w \in S \text{ and } \theta_v, \theta_w \in \mathbb{R}_{>0}\} \tag{7}$$

是包含 $S$ 的最小凸锥。

**引理 5.4**。 设 $C \subseteq \mathbb{R}^n$ 为一个集合。那么， $C$ 的凸性意味着对于任意 $p \in \mathbb{R}^n$，锥$_p$（$C$）是凸的。

证明。 当 $p \in C$ 时，由注 2.2 可知结论成立。其余情况，采用逆否命题证明。我们将证明，锥$_p$($C$) 的非凸性意味着 $C$ 非凸。由锥$_p$ ($C$) 的非凸性可知，存在两个不同的点 $v$， $w \in$ 锥$_p$ ($C$)，以及某个 $\theta \in (0,\ 1)$，使得$\theta v + (1 - \theta)w \notin$ 锥$_p$($C$)。

$v \in$ 锥$_p$($C$) 的含义是， $p + t_v v \in C$ 对某个 $t_v > 0$ 成立。类似地， $w \in$ 锥$_p$ ($C$) 意味着 $p + t_w w \in C$ 对某个 $t_w > 0$ 成立。现在令 $\tilde{v} = t_v v$ 和 $\tilde{w} = t_w w$。于是， $\tilde{v}$， $\tilde{w} \in$ 锥$_p$($C$)。令 $\theta = \dfrac{\theta t_w}{\theta t_w + (1-\theta) t_v}$，即可构造出 $\tilde{v}$ 和 $\tilde{w}$ 的凸组合：

$$\tilde{\theta}\tilde{v} + (1 - \tilde{\theta})\tilde{w},$$

这等于 $\dfrac{t_v t_w}{\theta t_w + (1-\theta) t_v}(\theta v + (1 - \theta)w)$。由于 $\theta v + (1 - \theta)w \notin$ 锥$_p$ ($C$)，且锥$_p$ ($C$)是一个锥，我们知道 $\tilde{\theta}\tilde{v} + (1 - \tilde{\theta})\tilde{w} \notin$ 锥$_p$ ($C$)。这意味着对于任意 $t > 0$，

$$p + t\,(\tilde{\theta}\tilde{v} + (1 - \tilde{\theta})\tilde{w}) \notin C.$$

对于 $t = 1$，上述内容等价于

$$\tilde{\theta}(p + \tilde{v}) + (1 - \tilde{\theta})(p + \tilde{w}) \notin C.$$

但 $p + \tilde{v}$ 和 $p + \tilde{w}$ 都属于 $C$，因为 $\tilde{v}$， $\tilde{w} \in$ 锥$_p$ ($C$)（因为 $v$， $w \in$ 锥$_p$ ($C$)）。这意味着 $C$不是凸集。■

**引理 5.5.** 真凸集 $C \subset \mathbb{R}^n$ 的补集的内部非空。

证明。 假设 $C^c$ 的内部为空。对于任意点 $v \in C^c$，必定存在一个 $w \in \mathbb{R}^n$，使得 $v + w \in C$ 且 $v - w \in C$。如果不存在这样的 $w$，那就意味着对于周围所有的向量 $w$，包括以 $v$ 为中心、半径任意的开球 $\epsilon$（即 $\text{ball}_v(\epsilon) := \{w \mid \|w\| \leq \epsilon\}$），都有 $v + w \in C^c$；这将与我们关于 $C^c$内部为空的假设矛盾。

既然存在这样的向量 $w$，使得 $v + w \in C$ 和 $v - w \in C$，那么很容易看出， $C$ 的凸性遭到了破坏，因为 $v = \frac{1}{2}$（$v + w$）$+ \frac{1}{2}$（$\bar{v} - w$）不属于 $C$。因此， $C^c$ 的内部不可能为空。■

**引理 5.6**。 每个真[3] 凸锥 $K \neq \mathbb{R}^n$ 都是某个半空间的子集。

证明。 根据引理 5.5，我们知道 $K^c$ 有非空内部。如果存在一个向量 $v \in K^c$，使得 $-v \in K^c$也成立，那么就可以取 $\{v\}$ 与 $K$的锥包。经过这一步后， $-v$ 不可能属于锥包($\{v\} \cup K$)。因此，除非不存在满足 $-v \in K^c$的向量$v \in K^c$ ，否则总能扩张凸锥 $K$ 。

由于对于 $K^c$ 中的每个向量，其相反向量都在 $K$ 中，因此 $-K^c$ 与 int ($K$)相同，这意味着 $K^c$ 本身现在是一个凸锥。因此，边界 cl ($K$) $\cap$ cl ($K^c$)是两个凸锥的交集，故为凸集。

边界，也就是点集 $\{v \in K | -v \in K\}$ ，除了作为凸锥而对加法封闭外，对数乘作用也封闭。这意味着边界是一个 ($n-1$) 维向量子空间，即一个超平面。

■

[3]一个是 $\mathbb{R}^n$真子集的锥。