

## Lecture 13 仿射几何、行列式 — 2025.12.09

教授：邵美悦

Scribe: 路人甲

## 目录

1 仿射变换	1
2 行列式	4
3 解析几何常用几何对象再研究	7

## 1 仿射变换

**仿射变换的表示** 映射是我们研究几何对象的重要工具。为了更好地研究仿射空间中的几何对象，我们有必要引入仿射变换的概念。在线性空间中，线性变换是保持向量加法和数乘的映射。向量空间对加法和数乘封闭，因此线性变换抓住了向量空间最核心的结构。我们在构造仿射空间上的映射时，自然也希望它们能够保持仿射空间最核心的结构——仿射组合。

**定义 1.1** (仿射变换). 若  $F: \mathbb{R}^n \rightarrow \mathbb{R}^n$  满足对任何  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$  以及  $\alpha \in \mathbb{R}$  都有

$$F(\mathbf{x}\alpha + \mathbf{y}\beta) = F(\mathbf{x})\alpha + F(\mathbf{y})\beta,$$

其中  $\alpha + \beta = 1$ ，则称  $F$  为一个仿射变换。

接下来，我们就来探讨仿射变换的一些基本性质。

首先，由定义 1.1 我们知道，仿射变换把点映射成点。这一性质看似寻常，但它实际上是仿射变换的一个重要特征。因为仿射空间中的点并不具备向量空间中的零元素，所以仿射变换必须确保点的映射仍然是点。

我们还知道，点可以看成是点加上一个向量。因此，仿射变换实际上可以看成是一个线性变换加上一个平移——这也就是说，假如我们对映射  $F$  做一个位移，那么它就变成了一个线性变换。

**定理 1.2.** 设  $F: \mathbb{R}^n \rightarrow \mathbb{R}^n$  是一个仿射变换, 那么

$$F(\mathbf{x}) = F(\mathbf{0}) + T(\mathbf{x}),$$

其中  $T: \mathbb{R}^n \rightarrow \mathbb{R}^n$  是一个线性变换。

既然线性变换可以用矩阵来表示, 仿射变换也一定能用矩阵来表示: 如果我们设  $T(\cdot)$  的表示矩阵为  $\mathbf{A}$ ,  $F(\mathbf{0}) = \mathbf{b}$ , 那么我们就有

$$F(\mathbf{x}) = \mathbf{Ax} + \mathbf{b}.$$

这就给了我们一个非常方便的工具来处理仿射变换。当然, 这种表示法也有其弊端。我们知道, 线性变换的复合是可以直接用矩阵乘法来表示的。但是, 如果我们采用这种表示法, 那么仿射变换的复合就不能简单地用矩阵乘法来表示了: 如果设  $F(\mathbf{x}) = \mathbf{Ax} + \mathbf{b}$ ,  $G(\mathbf{x}) = \mathbf{Cx} + \mathbf{d}$ , 那么

$$G \circ F(\mathbf{x}) = \mathbf{CAx} + \mathbf{Cb} + \mathbf{d}.$$

这就比较让人难过了: 它的形式实在不太美观。为了能够减轻仿射变换复合时产生的痛苦, 我们需要一种新的表示法。既然线性变换的复合能用矩阵乘法来表示, 我们自然也希望仿射变换的复合也能用矩阵乘法来表示。因此, 我们一定要把仿射变换表示成一个表示矩阵乘上一个坐标向量的形式。这并不是什么非常天才的想法, 想必大家都能想到, 仿射变换  $\mathbf{y} = F(\mathbf{x}) = \mathbf{Ax} + \mathbf{b}$  可以等价地表示为

$$\begin{bmatrix} \mathbf{y} \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{0} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ 1 \end{bmatrix}$$

这样一来, 仿射变换的复合就可以用矩阵乘法来表示了。顺道一提, 矩阵

$$\begin{bmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{0} & 1 \end{bmatrix}$$

被称为仿射变换矩阵, 向量  $[\mathbf{x}^T, 1]^T$  被称为齐次坐标。

从几何的角度, 这种变换也有其意义。普通的矩阵-向量乘法总是将原点映射到原点, 因此无法呈现平移; 为了解决这个问题, 我们在所有向量上扩增一个坐标 1, 将原空间映射到一个更高维的空间的一个子集以进行变换。在该空间中, 原空间占据了前  $n$  个坐标, 而最后一个坐标恒为 1。于是, 原空间的原点被映射到了  $[0, \dots, 0, 1]^T \in \mathbb{R}^{n+1}$ , 原空间的平移便可借由高维空间的线性变换来实现 (即为错切变换)。

接下来, 我们来看看一些常用的仿射变换及其对应的仿射变换矩阵。这里我直接引用 Snape 学长的笔记, 如图 1 所示。

仿射变换改变了原空间的坐标系, 但并不改变空间本身的结构。因此, 在仿射变换下, 一些几何性质是保持不变的。例如,

**TABLE 2.3**  
Affine  
transformations  
based on  
Eq. (2-45).

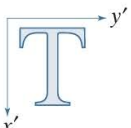
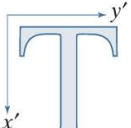
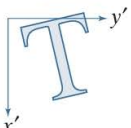
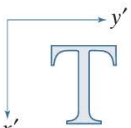
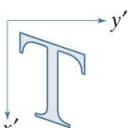
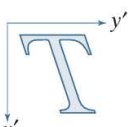
Transformation Name	Affine Matrix, A	Coordinate Equations	Example
Identity	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$x' = x$ $y' = y$	
Scaling/Reflection (For reflection, set one scaling factor to -1 and the other to 0)	$\begin{bmatrix} c_x & 0 & 0 \\ 0 & c_y & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$x' = c_x x$ $y' = c_y y$	
Rotation (about the origin)	$\begin{bmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$x' = x \cos \theta - y \sin \theta$ $y' = x \sin \theta + y \cos \theta$	
Translation	$\begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix}$	$x' = x + t_x$ $y' = y + t_y$	
Shear (vertical)	$\begin{bmatrix} 1 & s_v & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$x' = x + s_v y$ $y' = y$	
Shear (horizontal)	$\begin{bmatrix} 1 & 0 & 0 \\ s_h & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$x' = x$ $y' = s_h x + y$	

图 1: 常用仿射变换及其对应的仿射变换矩阵

1. 点之间的共线关系：共线点在变换后仍然共线。
2. 直线之间的平行关系：平行直线在变换后仍然平行。
3. 集合的凸性：凸集在变换后仍然是凸集。
4. 平行线段的比例关系：平行线段的比例在变换后保持不变，即  $\|\overrightarrow{AB}\|_2 : \|\overrightarrow{CD}\|_2 = \|\overrightarrow{F(A)F(B)}\|_2 : \|\overrightarrow{F(C)F(D)}\|_2$ 。
5. 质心的位置：质心在变换后仍然是质心。

上面这些结论都很好验证，我们不再赘述。其中第四条性质尤为重要。我们知道，共线线段的比例同时也是三角形面积的比例，在更高维的空间中也是体积的比例。因此，仿射变换保持了共底面、共顶点的锥体的体积的比例关系。

## 2 行列式

我们刚刚提到，仿射变换保持了共底面、共顶点的锥体的体积的比例关系。这说明，在仿射变换下，这些锥体被同时放大或缩小了相同的比例。那么这个比例到底是多少呢？从直觉上我们知道，平移变换不会改变体积，因此改变体积的关键在于线性变换。于是，核心问题就变为：线性变换  $\mathbf{A}$  到底把图形的大小改变了多少。

体积是一个比较复杂的概念，它涉及到了积分等内容。我们不妨先从最简单的情形入手：单位立方体的体积。我们知道，标准正交基会被映射成  $\mathbf{A}$  的各个列向量——因此，单位立方体会被映射成由  $\mathbf{A}$  的列向量张成的平行六面体。于是，问题就变成了：由  $\mathbf{A}$  的列向量张成的平行六面体的体积是多少。我们不妨记这个体积为  $\det(\mathbf{A})$ 。接下来，我们就来研究这个量满足哪些性质，进而推导出它的计算方法。

首先，单位立方体的体积为 1，因此我们有  $\det(\mathbf{I}) = 1$ 。

其次，我们知道，交换两个列向量会改变平行六面体的方向，从而改变体积的符号。因此，我们有：

$$\det([\mathbf{a}_1, \dots, \mathbf{a}_i, \dots, \mathbf{a}_j, \dots, \mathbf{a}_n]) = -\det([\mathbf{a}_1, \dots, \mathbf{a}_j, \dots, \mathbf{a}_i, \dots, \mathbf{a}_n]).$$

再次，如果某个列向量是另外两个向量的线性组合，对应地体积也应该是另外两个体积的线性组合——这可以由割补法轻松得到。因此，我们有：

$$\det([\mathbf{a}_1, \dots, \mathbf{a}_i k_1 + \mathbf{b}_i k_2, \dots, \mathbf{a}_n]) = k_1 \det([\mathbf{a}_1, \dots, \mathbf{a}_i, \dots, \mathbf{a}_n]) + k_2 \det([\mathbf{a}_1, \dots, \mathbf{b}_i, \dots, \mathbf{a}_n]).$$

满足这三条性质的函数  $\det(\cdot)$  被称为行列式函数。我们可以证明：满足以上三条性质的行列式函数唯一。

**定理 2.1.** 行列式是使得  $\det(\mathbf{I}) = 1$  的唯一交错多线性泛函。

证明. 在这个证明中, 我们将给出行列式的计算方法——由此证明唯一性。注意到, 矩阵  $\mathbf{A}$  的每一个列向量都可以写成

$$\mathbf{a}_i = \sum_{j=1}^n \mathbf{e}_j a_{j,i},$$

其中  $\mathbf{e}_j$  是标准正交基。因此, 我们有

$$\begin{aligned} \det(\mathbf{A}) &= \det \left( \left[ \sum_{i_1=1}^n \mathbf{e}_{i_1} a_{i_1,1}, \sum_{i_2=1}^n \mathbf{e}_{i_2} a_{i_2,2}, \dots, \sum_{i_n=1}^n \mathbf{e}_{i_n} a_{i_n,n} \right] \right) \\ &= \sum_{i_1=1}^n \sum_{i_2=1}^n \cdots \sum_{i_n=1}^n \det([\mathbf{e}_{i_1}, \mathbf{e}_{i_2}, \dots, \mathbf{e}_{i_n}]) a_{i_1,1} \cdots a_{i_n,n} \\ &= \sum_{(i_1, i_2, \dots, i_n) \in S(1, 2, \dots, n)} \det([\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n]) (-1)^{s(i_1, \dots, i_n)} a_{i_1,1} \cdots a_{i_n,n} \quad \text{其中 } S \text{ 表示排列,} \\ &\hspace{25em} s \text{ 表示逆序数} \\ &= \sum_{(i_1, i_2, \dots, i_n) \in S(1, 2, \dots, n)} (-1)^{s(i_1, \dots, i_n)} a_{i_1,1} a_{i_2,2} \cdots a_{i_n,n}. \end{aligned}$$

第二个等号利用了求和的可交换性以及行列式的多线性性质, 第三个等号利用了行列式的交错性, 第四个等号利用了  $\det(\mathbf{I}) = 1$ .

这就给出了行列式的计算方法。由此可见, 行列式是唯一的。  $\square$

当然, 这个证法不够优雅。更优雅的证明方法是: 利用初等变换将矩阵化为单位阵, 然后利用行列式在初等变换下的变化规律来计算行列式的值。这种方法更符合线性代数的思维方式, 但鉴于同学们对此比较熟悉, 在此不再赘述。

既然我们知道了单位立方体在变换  $\mathbf{A}$  下的体积是  $\det(\mathbf{A})$ , 那么任意图形在变换  $\mathbf{A}$  下的体积又是多少呢? 我们知道,  $\mathbb{R}^n$  空间中的任意图形的体积都可以表示为  $\int_{\Omega} d\mathbf{x}$ , 而变换后的体积为  $\int_{\mathbf{A}(\Omega)} d\mathbf{A}(\mathbf{x})$ . 这里,  $d\mathbf{x}$  和  $d\mathbf{A}(\mathbf{x})$  分别表示在原空间和变换后空间中的体积微元。显然, 体积微元应满足

$$d\mathbf{A}(\mathbf{x}) = |\det(\mathbf{A})| d\mathbf{x}.$$

因为它们二者分别是小平行六面体和小立方体的体积, 而小平行六面体的体积是小立方体体积的  $|\det(\mathbf{A})|$  倍。因此, 我们有

$$\int_{\mathbf{A}(\Omega)} d\mathbf{A}(\mathbf{x}) = \int_{\Omega} |\det(\mathbf{A})| d\mathbf{x}.$$

亦即任意图形在变换  $\mathbf{A}$  下的体积是原体积的  $|\det(\mathbf{A})|$  倍。于是, 我们证明了行列式  $\det(\mathbf{A})$  是线性变换的体积放缩因子。

从行列式的几何意义出发，我们还可以理解 Cramer 法则为什么成立。如图 2 所示，线性方程组  $\mathbf{Ax} = \mathbf{b}$  的解显然是  $\mathbf{x} = [1, 1/2]^T$ 。另一方面，在图中我们可以看到，由  $\mathbf{a}_1, \mathbf{b}$  构成的蓝色平行四边形与由  $\mathbf{a}_1, \mathbf{a}_2$  构成的紫色平行四边形的面积比，正好是  $\mathbf{b}$  在  $\mathbf{a}_2$  方向上的投影，也就是解的第二个分量  $1/2$ ；由  $\mathbf{b}, \mathbf{a}_2$  构成的橘黄色平行四边形与由  $\mathbf{a}_1, \mathbf{a}_2$  构成的紫色平行四边形的面积比，正好是  $\mathbf{b}$  在  $\mathbf{a}_1$  方向上的投影，也就是解的第一个分量  $1$ 。这正是 Cramer 法则。

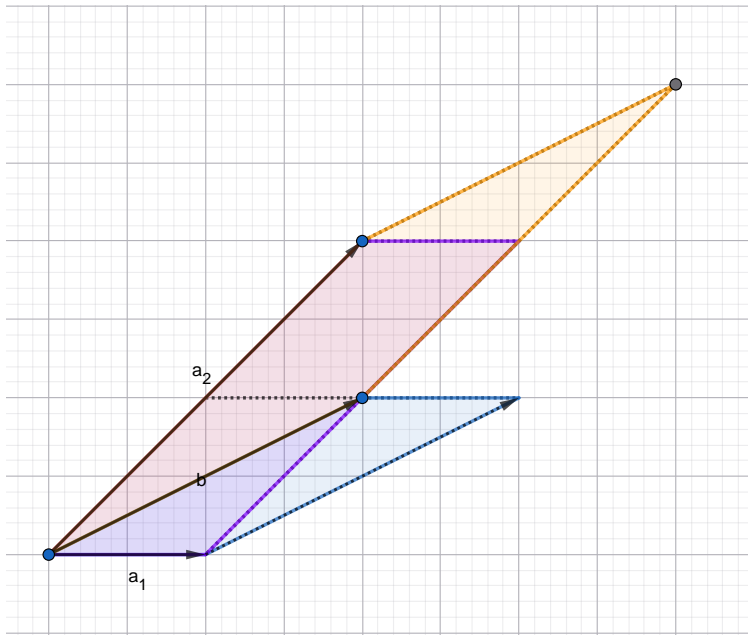


图 2: Cramer 法则的几何意义

严格的证明大家大一都学习过，故在此不做解释，仅给出大体思路：

$$\det \left( \left[ \mathbf{a}_1, \dots, \mathbf{a}_{i-1}, \sum_{k=1}^n \mathbf{a}_k x_k, \mathbf{a}_{i+1}, \dots, \mathbf{a}_n \right] \right) / \det(\mathbf{A}) = x_i.$$

下面我们接着来看行列式在几何中的应用。假如平面内三个点  $A, B, C$  构成一个三角形，那么平面内任意一点  $P$  都可以表示为  $A, B, C$  的仿射组合。那么，组合的系数又是什么呢？我们不妨设  $P = Au + Bv + Cw$ ，其中  $u + v + w = 1$ 。既然这是仿射组合，受仿射变换的矩阵表示的启发，我们不妨考虑将四个点  $A, B, C, P$  都升一维，写成齐次坐标的形式，于是我们有

$$\begin{bmatrix} P_x \\ P_y \\ 1 \end{bmatrix} = \begin{bmatrix} A_x & B_x & C_x \\ A_y & B_y & C_y \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} u \\ v \\ w \end{bmatrix}.$$

我们把  $[u, v, w]^T$  称为  $P$  关于  $A, B, C$  的重心坐标。接下来，我们来计算这个重心坐标。利用

Cramer 法则，我们有

$$\begin{aligned} u &= \frac{\det([\tilde{\mathbf{P}}, \tilde{\mathbf{B}}, \tilde{\mathbf{C}}])}{\det([\tilde{\mathbf{A}}, \tilde{\mathbf{B}}, \tilde{\mathbf{C}}])} = \frac{S_{\triangle PBC}}{S_{\triangle ABC}} \\ v &= \frac{\det([\tilde{\mathbf{A}}, \tilde{\mathbf{P}}, \tilde{\mathbf{C}}])}{\det([\tilde{\mathbf{A}}, \tilde{\mathbf{B}}, \tilde{\mathbf{C}}])} = \frac{S_{\triangle PAC}}{S_{\triangle ABC}} \\ w &= \frac{\det([\tilde{\mathbf{A}}, \tilde{\mathbf{B}}, \tilde{\mathbf{P}}])}{\det([\tilde{\mathbf{A}}, \tilde{\mathbf{B}}, \tilde{\mathbf{C}}])} = \frac{S_{\triangle PAB}}{S_{\triangle ABC}}, \end{aligned}$$

利用此关系，我们可以轻松地写出重心、内心、外心、垂心等重要点的坐标表达式，还可以轻松地得到如 Ceva 定理、Menelaus 定理等平面几何中的重要结论，在此不再赘述。在数值逼近中，重心坐标也有重要的应用，例如高维插值等。如图 3，我们只知道六个点  $A, B, C, D, E, F$  的函数值  $f(A), f(B), f(C), f(D), f(E), f(F)$ ，但是我们可以大胆地假设函数  $f$  是仿射变换，那么我们就可以利用重心坐标来对任意点  $P$  处的函数值进行插值，进而得到整个区域内的函数值分布——依次方法，我们生成了一个连续的函数。

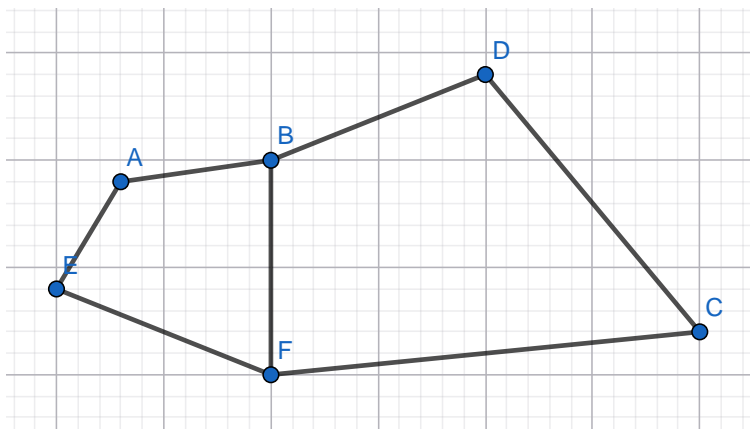


图 3: 利用重心坐标进行插值

### 3 解析几何常用几何对象再研究

本节知识点比较细碎，本人能力有限，无法以一个统一的逻辑将它们串联起来，望见谅。

我们知道，仿射空间  $\mathcal{A}$  可以表示成  $\mathcal{A} = \mathbf{v} + \mathcal{V}$ ，其中  $\mathbf{v} \in \mathcal{V}$  是一个点。显然  $\dim \mathcal{A} = \dim \mathcal{V}$ 。当  $\dim \mathcal{A} = 1$  时， $\mathcal{A}$  是一条直线： $\mathcal{A} = \{\mathbf{v} + \mathbf{w}t : t \in \mathbb{R}\}$ 。熟悉解析几何的同学应该知道，这是直线的参数方程。所以把直线定义成参数方程形式是最自然的。当  $\dim \mathcal{A} = n - 1$  时，我们称  $\mathcal{A}$  为超平面。我们还知道，超平面可以表示成

$$A_1x_1 + A_2x_2 + \cdots + A_nx_n = z,$$

如果我们表示成矩阵形式，就是

$$\begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_n \end{bmatrix}^T \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = z.$$

这说明，超平面上的点与一个给定向量的内积为定值——这不正是法向量的定义吗？因此，超平面的法向量正是  $\mathbf{A} = [A_1, \dots, A_n]^T$ .

**凸集分离定理** 从直观上，我们知道，如果两个凸集不相交，那么它们之间一定存在一个超平面将它们分开。这个结论被称为凸集分离定理。如图 4 所示。

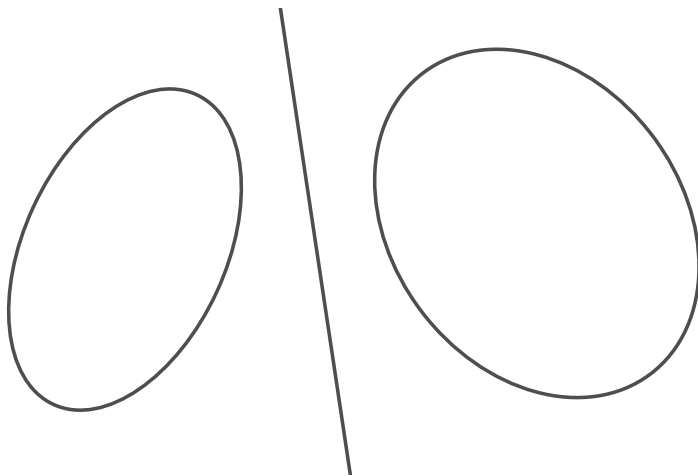


图 4: 凸集分离定理

为了证明这个结论，我们不妨考虑最简单的情形：其中一个凸集退化成一个点。此时的凸集分离定理叫做凸集支撑定理。

**定理 3.1 (凸集支撑定理).** 若  $\Omega \subset \mathbb{R}^n$  是非空凸集， $\mathbf{x}_0 \in \mathbb{R}^n \setminus \Omega$ ，则存在超平面  $\mathcal{P}$  使得  $\mathbf{x}$  和  $\Omega$  被  $\mathcal{P}$  分离，即存在  $\boldsymbol{\omega} \in \mathbb{R}^n \setminus \Omega$ ， $\alpha \in \mathbb{R}$  使得对一切  $\mathbf{y} \in \Omega$  有

$$(\boldsymbol{\omega}^T \mathbf{x}_0 + \alpha)(\boldsymbol{\omega}^T \mathbf{y} + \alpha) \leq 0.$$

结论有时也写作

$$\sup_{\mathbf{y} \in \Omega} \boldsymbol{\omega}^T \mathbf{y} \leq \boldsymbol{\omega}^T \mathbf{x}_0.$$

证明. 定理要证的内容我们可能一时半会不太看得懂，我们不妨先分析一下它的含义。我们知道，超平面可以表示成

$$\boldsymbol{\omega}^T \mathbf{z} + \alpha = 0.$$



其中  $\omega$  是超平面的法向量。于是，定理要证的内容实际上是说，存在一个超平面  $\mathcal{P}$  使得对一切  $y \in \Omega$  都有  $x_0$  与  $y$  分居超平面的两侧，如图 5 所示。

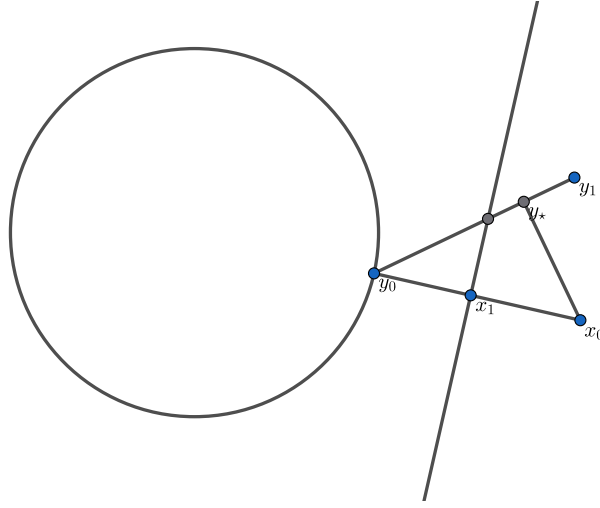


图 5: 凸集支撑定理示意图

既然  $x_0 \notin \Omega$ ，我们不妨设

$$y_0 = \arg \min_{y \in \Omega} \|y - x_0\|_2 \geq 0.$$

也就是说， $y_0$  是  $\Omega$  中距离  $x_0$  最近的点。由于  $\Omega$  是凸集，因此这个最小值一定存在。接下来，我们来构造超平面  $\mathcal{P}$ 。我们不妨设

$$\omega = y_0 - x_0.$$

然后，我们在  $\omega$  上作一个垂直于  $\omega$  的超平面  $\mathcal{P}$ ，并设  $\mathcal{P} \cap \omega = x_1$ 。于是

$$\omega^T(x_0 - x_1) < 0$$

$$\omega^T(y_0 - x_1) > 0.$$

这两个不等式是显然成立的：如图 5， $x_0 - x_1$  与  $\omega$  反向，而  $y_0 - x_1$  与  $\omega$  同向。

接下来，我们只需证  $\Omega$  中的所有点都在  $\mathcal{P}$  的同一侧，即  $\forall y \in \Omega$ ，有  $\omega^T(y - x_1) > 0$ 。我们用反证法。假设  $\exists y_1 \in \Omega$ , s.t.  $\omega^T(y_1 - x_1) < 0$ 。这也就是说， $y_1$  与  $x_0$  同侧。现在，我们只需说明这个条件与  $y_0 = \arg \min_{y \in \Omega} \|x_0 - y\|$  矛盾。

如图 5，如果存在  $y_1$  与  $x_0$  位于同侧，那么直觉告诉我们  $y_0$  不可能是离  $x_0$  最近的点：垂线段最短，我们一定能找到一个  $y_*$  使得  $y_* = \arg \min_{y \in \Omega} \|x_0 - y\|$ 。我们下面的证明思路就将围绕着这个直觉展开。

我们取  $y = \alpha y_1 + (1 - \alpha)y_0$  ( $\alpha \in [0, 1]$ )。由于凸集对凸组合封闭，有  $y \in \Omega$ 。下面我们证明存在一

个  $\alpha$  使得  $\|\mathbf{y} - \mathbf{x}_0\| < \|\mathbf{y}_0 - \mathbf{x}_0\|$ . 不妨考虑

$$\begin{aligned}
D(\alpha) &= \|\mathbf{x}_0 - \mathbf{y}\|^2 \\
&= \|\mathbf{x}_0 - (\alpha\mathbf{y}_1 + (1 - \alpha)\mathbf{y}_0)\|^2 \\
&= \|\mathbf{x}_0 - \mathbf{y}_0 + \alpha(\mathbf{y}_0 - \mathbf{y}_1)\|^2 \\
&= \|\mathbf{x}_0 - \mathbf{y}_0\|^2 + 2\alpha(\mathbf{x}_0 - \mathbf{y}_0)^T(\mathbf{y}_0 - \mathbf{y}_1) + \alpha^2\|\mathbf{y}_0 - \mathbf{y}_1\|^2.
\end{aligned}$$

这是一个关于  $\alpha$  的二次函数, 满足  $D(0) = \|\mathbf{x}_0 - \mathbf{y}_0\|^2$ . 它的对称轴是

$$\begin{aligned}
\alpha &= \frac{-2(\mathbf{x}_0 - \mathbf{y}_0)^T(\mathbf{y}_0 - \mathbf{y}_1)}{2\|\mathbf{y}_0 - \mathbf{y}_1\|^2} \\
&= \frac{2\boldsymbol{\omega}^T(\mathbf{y}_0 - \mathbf{x}_1 - (\mathbf{y}_1 - \mathbf{x}_1))}{2\|\mathbf{y}_0 - \mathbf{y}_1\|^2} \\
&= \frac{2(\boldsymbol{\omega}^T(\mathbf{y}_0 - \mathbf{x}_1) - \boldsymbol{\omega}^T(\mathbf{y}_1 - \mathbf{x}_1))}{2\|\mathbf{y}_0 - \mathbf{y}_1\|^2} \\
&> 0.
\end{aligned}$$

因此,  $\exists \alpha > 0$ , 使得  $D(\alpha) < D(0)$ . 这就完成了证明。  $\square$

由凸集支撑定理, 我们可以轻松地证明凸集分离定理。

证明. 设  $\mathbf{x}_0 \in \Omega_1$ , 则由凸集支撑定理, 存在超平面  $\mathcal{P}$  使得对一切  $\mathbf{y} \in \Omega_2$  有

$$(\boldsymbol{\omega}^T \mathbf{x}_0 + \alpha)(\boldsymbol{\omega}^T \mathbf{y} + \alpha) \leq 0.$$

由于  $\mathbf{x}_0 \in \Omega_1$ , 因此对一切  $\mathbf{x} \in \Omega_1$  都有

$$(\boldsymbol{\omega}^T \mathbf{x} + \alpha)(\boldsymbol{\omega}^T \mathbf{x}_0 + \alpha) \geq 0.$$

综上, 对任意  $\mathbf{x} \in \Omega_1$ ,  $\mathbf{y} \in \Omega_2$  都有

$$(\boldsymbol{\omega}^T \mathbf{x} + \alpha)(\boldsymbol{\omega}^T \mathbf{y} + \alpha) \leq 0.$$

这就完成了证明。  $\square$

**锥** 在解析几何中, 锥体通常被定义为由一个顶点和一个底面构成的几何体。然而, 从仿射几何的角度来看, 锥体可以被更一般地定义为由一个点集通过仿射组合生成的集合。

**定义 3.2** (锥). 若  $\mathcal{K} \subset \mathbb{R}^n$ , 对任何  $\mathbf{x} \in \mathcal{K}$  以及  $\alpha \in (0, +\infty)$  都有  $\mathbf{x}\alpha \in \mathcal{K}$ , 则称  $\mathcal{K}$  为一个锥。

**例 3.3.** 集合  $\{\mathbf{H} : \mathbf{H}^* = \mathbf{H} \succcurlyeq (\succeq) 0\}$  构成一个凸锥。其中, *Hermite* 半正定矩阵构成的集合是闭凸锥。

现在，我们来研究一下分离定理在凸锥上的应用。设  $\Omega$  是一个凸锥， $\mathbf{x}_0 \in \mathbb{R}^n \setminus \Omega$ 。由凸集支撑定理，存在法向量  $\boldsymbol{\omega}$  使得

$$\sup_{\mathbf{y} \in \Omega} \boldsymbol{\omega}^T \mathbf{y} \leq \boldsymbol{\omega}^T \mathbf{x}_0.$$

由锥的定义，若  $\mathbf{y} \in \Omega$ ，有  $\mathbf{y}\lambda \in \Omega, \forall \lambda \in (0, +\infty)$ 。因此，若  $\boldsymbol{\omega}^T \mathbf{y} > 0$ ，我们取  $\lambda \rightarrow +\infty$ ，则  $\boldsymbol{\omega}^T(\mathbf{y}\lambda) \rightarrow +\infty$ ，不可能一直小于固定值  $\boldsymbol{\omega}^T \mathbf{x}_0$ 。这就产生了一个矛盾！

因此，对于凸锥，分离不等式要成立，必须有

$$\boldsymbol{\omega}^T \mathbf{y} \leq 0, \quad \forall \mathbf{y} \in \Omega.$$

当然，如果我们考虑反向的分离，不等号方向就会相反。不论怎样，凸锥都可以被一个过原点的超平面分离出来。分离定理在凸锥上的这个特性在优化理论中有重要的应用。

现在，我们考察所有能“从一侧观察”这个锥的向量：

**定义 3.4** (对偶锥). 锥  $\mathcal{K}$  的对偶锥  $\mathcal{K}^*$  定义为

$$\mathcal{K}^* = \{\mathbf{y} \in \mathbb{R}^n : \mathbf{y}^T \mathbf{x} \geq 0, \forall \mathbf{x} \in \mathcal{K}\},$$

即  $\mathcal{K}^*$  由与  $\mathcal{K}$  中的所有向量夹角小于  $\pi/2$  的向量组成。

显然，若  $\mathcal{K}$  是闭凸锥，则有  $\mathcal{K}^{**} = \mathcal{K}$ ，即对偶锥的对偶锥是原锥。

**例 3.5.** 考虑最简单的凸锥： $\mathbb{R}_+^n$ ——它的对偶锥是它本身。这种自对偶性非常特殊且有用。

下面我们来看看锥这个概念在线性代数中的应用。

**例 3.6.** 设  $\mathbf{A}^T = \mathbf{A} \succ 0$ 。证明：存在  $\mathbf{x} > 0$ ，使得  $\mathbf{A}\mathbf{x} > 0$ 。

注：这里的“ $> 0$ ”指的是所有分量都是正数。

证明. 考虑  $\Omega_1 = \{\mathbf{x} > 0\}, \Omega_2 = \{\mathbf{x} : \mathbf{A}\mathbf{x} > 0\}$ 。我们只需证  $\Omega_1 \cap \Omega_2 \neq \emptyset$ 。不妨用反证法。假设  $\Omega_1 \cap \Omega_2 = \emptyset$ ，则由凸锥分离定理， $\exists \boldsymbol{\omega} \geq 0$ ，有

$$\boldsymbol{\omega}^T \mathbf{x} \geq 0 \geq \boldsymbol{\omega}^T \mathbf{y}.$$

对任意  $\mathbf{z} > 0$ ，取  $\mathbf{y} = \mathbf{A}\mathbf{z} > 0$ ，则有

$$\boldsymbol{\omega}^T \mathbf{A}\mathbf{z} \leq 0.$$

由于  $\mathbf{z}$  是任意的正向量，我们令  $\mathbf{z} \rightarrow \boldsymbol{\omega}$ ，则有

$$\boldsymbol{\omega}^T \mathbf{A}\boldsymbol{\omega} \leq 0,$$

这与  $\mathbf{A} \succ 0$  矛盾。因此，假设不成立，结论得证。 □