



支持向量机(SVM)数学基础

基础几何概念回顾

我们从初中数学开始，给定一个二维欧氏空间的直线方程：

$$l : ax + by + c = 0$$

其中系数 (a, b, c) 唯一确定直线位置。给定平面上的点 $p(x_0, y_0)$ ，其到直线的有向距离公式为：

$$d = \frac{|ax_0 + by_0 + c|}{\sqrt{a^2 + b^2}}$$

现代数学视角：这个方程定义了二维空间中的一维仿射子空间：

$$\mathbb{A}^1 = \{(x, y) \in \mathbb{R}^2 \mid ax + by + c = 0\}$$

该子空间具有以下特性：

1. 平移不变性（仿射空间核心性质）
2. 不存在原点（与线性子空间的核心区别）
3. 在微分几何中可视为平凡的一维流形

线性代数重构

为衔接机器学习表示惯例，我们进行参数替换：

$$\mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad b = c$$

此时直线方程转化为：

$$\mathbf{w}^\top \mathbf{x} + b = w_1 x_1 + w_2 x_2 + b = 0$$

核心几何量解析：

1. **法向量：** \mathbf{w} 正交于直线方向（图1示意）
2. **偏置项：** b 控制直线与原点的相对位置

3. **参数冗余性**：对任意 $\lambda \neq 0$ ，方程 $\lambda(\mathbf{w}^\top \mathbf{x} + b) = 0$ 表示同一超平面

点距公式的现代形式：

$$d(\mathbf{x}) = \frac{|\mathbf{w}^\top \mathbf{x} + b|}{\|\mathbf{w}\|}$$

公式内涵：

- 分子：代数距离（带符号的线性输出）
- 分母：法向量范数（几何归一化因子）
- 绝对值运算：保持距离度量的非负性

高维空间推广

将上述结论推广到 n 维空间，我们得到超平面方程：

$$\mathcal{H} : \mathbf{w}^\top \mathbf{x} + b = 0 \quad (\mathbf{w} \in \mathbb{R}^n, \mathbf{x} \in \mathbb{R}^n)$$

几何性质继承：

1. 法向量 \mathbf{w} 始终正交于超平面切空间
2. 原点距离公式保持形式不变： $d(0) = \frac{|b|}{\|\mathbf{w}\|}$
3. 参数缩放等价性： $\forall \lambda > 0, (\lambda \mathbf{w}, \lambda b)$ 定义相同超平面

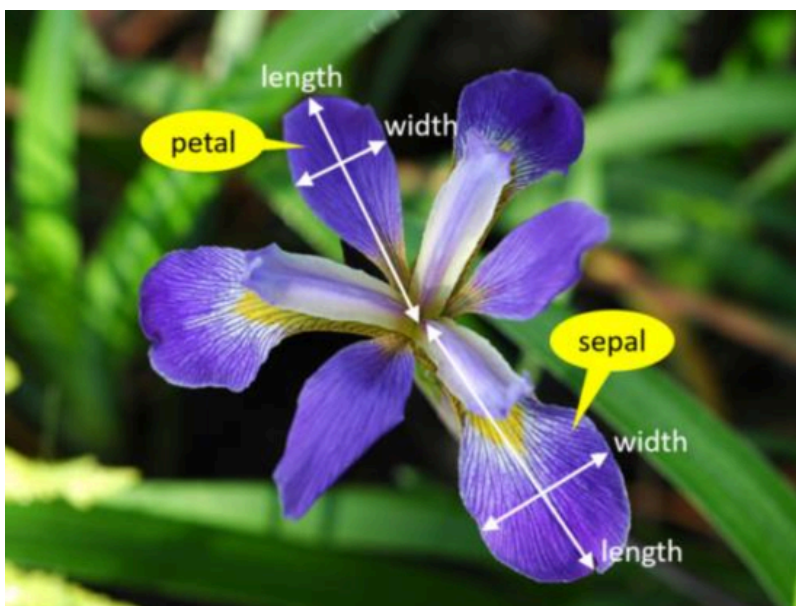
二分类问题

鸢尾花数据最初由美国植物学家 Edgar Anderson 在20世纪20年代收集，包含3种鸢尾花（山鸢尾、变色鸢尾、维吉尼亚鸢尾）的萼片和花瓣尺寸数据（各50个样本）。

在1936年的论文中，Fisher以该数据集为例，证明了多变量（多个特征）统计方法在分类问题中的有效性。他通过计算不同特征的线性组合（即判别函数），成功将三类鸢尾花分离，成为统计学史上首个高维（多个特征）数据分类的范例。

为了理解前辈的工作，我们用Fisher的方法重现这项工作。为了方便教学，我们暂时只考虑两种鸢尾花分类问题。仅保留山鸢尾和变色鸢尾各50个样本，删掉第三类50个数据。

要做的事情是这样，100个样本数据被 Anderson 定义了4个特征，分别为萼片(sepal)长度，宽度；花瓣(petal)长度，宽度。



并且他根据这四个特征综合判断，一株鸢尾花应该属于山鸢尾还是变色鸢尾，并将结果标注为：0表示山鸢尾，1表示变色鸢尾。因此这是一个经权威专家标注过的分类问题，现在成了监督学习的经典案例。接下来我们就尝试用支持向量机(SVM)来解决这个问题。

注意，SVM只是整个机器学习中的一小段，因为喂给机器的数据，是已经提取了4个特征的样本数据。实际上的机器学习，要自己处理原始数据，用PCA，LDA等方法，提取出有用的特征，然后再进入SVM环节。

线性 SVM 的优化问题

SVM 的目标是找到一个超平面（即找到参数 w, b ）：

$$w^T x + b = 0$$

使得数据点尽可能被正确分类，并且间隔（margin）最大。即在约束条件下：

$$y_i(w^T x_i + b) \geq 1, \quad \forall i$$

实现最优化目标：

$$\min_{w,b} \frac{1}{2} \|w\|^2$$

这就是 **硬间隔 SVM (hard-margin SVM)**，它假设数据完全线性可分。这是一个典型的凸二次规划问题，由于存在不等式约束，我们可以通过拉格朗日乘子法求解。

1. 复习高中数学

高中生经常做这种题：给定一个函数 $f(x)$, $x \in D \subseteq \mathbb{R}$, 其中 $D = [a, b]$, 求 $f(x)$ 在 D 上的最值。最值判定法则的标准套路就是三步走：

1. **求导数**，找到满足 $f'(x) = 0$ 的**驻点**，这些可能是局部极值点。
2. **检查边界点** $x = a$ 和 $x = b$ ，因为全局最值可能出现在边界。
3. **比较函数值**，确定全局最小值和最大值。

核心思想：

- 如果最优解在区间内部，则导数为零（类似于优化问题中梯度消失的情况）。
- 如果最优解在区间边界，则只能是边界的某个点（类似于约束优化中解落在边界）。

2. 复习本科数学

把上述问题推广到两个变量的函数：给定一个函数 $f(x, y)$, $(x, y) \in D \subseteq \mathbb{R}^2$, 求函数在 D 上的最值。到了二维及以上，事情就变得有点微妙了。首先，这个 D 就要分两种情况考虑：

一， D 是 \mathbb{R}^2 上的一条**封闭曲线** γ ，对应等式约束求解最值，例如：

目标函数：

$$f(x, y) = x^2 + 2x - y^2 + 4y$$

约束条件 γ ：

$$h(x, y) = \frac{x^2}{2^2} + \frac{y^2}{3^2} - 1 = 0$$

假设我们照猫画虎，第一步先求 $\nabla f = 0$ 的极值点：

$$\nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right) = (2x + 2, -2y + 4) = (0, 0)$$

解方程组得到极值点 $p(x, y)$:

$$\begin{cases} 2x + 2 = 0 & \Rightarrow x = -1 \\ -2y + 4 = 0 & \Rightarrow y = 2 \end{cases}$$

然而并没有什么卵用, 这个点 p 几乎不可能恰好落在约束曲线 γ 上。怎么办捏? 打不过就加入, 大神拉格朗日想出来个拉郎配的好办法, 把 γ 拉到 f 上, 变成一个新的函数 $\mathcal{L}(x, y, \lambda)$, 结合了原始目标函数和约束条件:

$$\mathcal{L}(x, y, \lambda) = f(x, y) + \lambda h(x, y)$$

这里, λ 是拉格朗日乘数, 它是一个待定系数, 通过它我们可以将约束条件融入到目标函数中。接下来就是新的标准套路三步走:

1. **计算偏导数**: 对拉格朗日函数 $\mathcal{L}(x, y, \lambda)$ 关于 x 、 y 和 λ 分别求偏导数, 并令这些偏导数等于零。这意味着:

$$\frac{\partial \mathcal{L}}{\partial x} = 0, \quad \frac{\partial \mathcal{L}}{\partial y} = 0, \quad \text{和} \quad \frac{\partial \mathcal{L}}{\partial \lambda} = 0$$

2. **解方程组**: 上述三个偏导数等于零的方程构成了一个方程组。解这个方程组可以找到 x 、 y 和 λ 的值, 其中 x 和 y 是满足约束条件下的可能最优解。
3. **验证解**: 确定哪些解是最优解。这通常涉及到二次导数检验或者直接比较候选点处的目标函数值来决定哪个是最大值或最小值。

接下来我们就用拉格朗日乘子法求解刚才搞不定的问题:

$$\begin{aligned} \mathcal{L}(x, y, \lambda) &= f(x, y) + \lambda h(x, y) \\ &= x^2 + 2x - y^2 + 4y + \lambda \left(\frac{x^2}{2^2} + \frac{y^2}{3^2} - 1 \right) \end{aligned}$$

求 $\nabla \mathcal{L} = 0$ 的解:

$$\begin{aligned} \nabla \mathcal{L} &= \left(\frac{\partial \mathcal{L}}{\partial x}, \frac{\partial \mathcal{L}}{\partial y}, \frac{\partial \mathcal{L}}{\partial \lambda} \right) \\ &= \left(2x + 2 + \frac{1}{2}\lambda x, -2y + 4 + \frac{2}{9}\lambda y, \frac{x^2}{2^2} + \frac{y^2}{3^2} - 1 \right) \\ &= (0, 0, 0) \end{aligned}$$

解下面的方程组（三个方程三个未知数，没毛病）：

$$2x + 2 + \frac{1}{2}\lambda x = 0 \quad (1)$$

$$-2y + 4 + \frac{2}{9}\lambda y = 0 \quad (2)$$

$$\frac{x^2}{2^2} + \frac{y^2}{3^2} - 1 = 0 \quad (3)$$

一共得到四组解（虽然由符号求解得出，但列出根式解无意义，我直接写成数值形式了）：

$$\begin{array}{llll} x : & -0.2102 & -0.59571 & -1.6466 & 1.8371 \\ y : & -2.9834 & 2.8638 & 1.7028 & 1.1860 \\ \lambda : & -15.0330 & -2.7147 & 1.5708 & 6.1774 \end{array}$$

分别代入这四组局部极值点，求出目标函数值：

$$\begin{array}{lll} x = -0.21016 & y = -2.9834 & f(x, y) = -21.2104 \\ x = -0.59571 & y = -2.8638 & f(x, y) = 2.4173 \\ x = -1.6466 & y = 1.7028 & f(x, y) = 3.3298 \\ x = 1.8371 & y = 1.1860 & f(x, y) = 10.3865 \end{array}$$

最终得到全局最大值和最小值分别为：

$$\begin{array}{lll} \min f : & -21.2103 & \text{at } (-0.2102, -2.9834) \\ \max f : & 3.3298 & \text{at } (-1.6466, 1.7028) \end{array}$$

用Python的scipy库中的minimize函数求解，得到的结果与上面一致。

二， D 是 \mathbb{R}^2 上的一个区域，对应不等式约束求解最值，记作：

$$\min f(x, y), \quad \text{s.t.} \quad g(x, y) \leq 0$$

注意，不等式约束，因为流程稍微复杂点，为防止低级错误，工程规范要求仅求最小值。如果优化目标是最大值，加个负号即可。

$$\begin{aligned}\text{objective : } f(x, y) &= x^2 + y^2 - 3x + 5y - 2 \\ \text{constraint : } g(x, y) &= \frac{x^2}{2^2} + \frac{y^2}{3^2} - 2 \leq 0\end{aligned}$$

1, 通过目标函数梯度为零 ($\nabla f = 0$), 寻找内部极值点:

$$\nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right) = (2x - 3, 2y + 5) = (0, 0)$$

解方程组得到极值点 $p_i(x_p, y_p)$ (可能没有解, 也可能有多个) :

$$\begin{cases} 2x_p - 3 = 0 & \Rightarrow x_p = 1.5 \\ 2y_p + 5 = 0 & \Rightarrow y_p = -2.5 \end{cases}$$

如果有多个解, 依次代入约束方程, 根据 $g(x_p, y_p) < 0$ 是否成立, 判断点 p_i 是否在**可行域**内:

$$\frac{x^2}{2^2} + \frac{y^2}{3^2} - 2 = \frac{1.5^2}{2^2} + \frac{(-2.5)^2}{3^2} - 2 \approx -0.743 < 0$$

将可行域外的解抛掉, 只保留可行域内的解暂存起来。

2, 通过拉格朗日函数梯度为零 ($\nabla \mathcal{L} = 0$), 寻找边界极值点:

$$\nabla \mathcal{L} = \left(\frac{\partial \mathcal{L}}{\partial x}, \frac{\partial \mathcal{L}}{\partial y}, \frac{\partial \mathcal{L}}{\partial \lambda} \right) = \left(\frac{\lambda x}{2} + 2x - 3, \frac{2\lambda y}{9} + 2y + 5, \frac{x^2}{4} + \frac{y^2}{9} - 2 \right) = (0, 0, 0)$$

得到一组代数方程:

$$\begin{cases} \frac{\lambda x}{2} + 2x - 3 = 0 \\ \frac{2\lambda y}{9} + 2y + 5 = 0 \\ \frac{x^2}{4} + \frac{y^2}{9} - 2 = 0 \end{cases}$$

解方程组得到极值点 $p_i(x_p, y_p)$ (依然可能没有解, 也可能有多个) :

$$\begin{array}{ll} x : & -0.575992 \quad 2.089986 \\ y : & 4.153736 \quad -2.858655 \\ \lambda : & -14.416810 \quad -1.129167 \end{array}$$

根据条件 $\lambda \geq 0$ 判断解是否有效，小于0则无效。因为：

$$\nabla f = -\lambda \nabla g$$

所以 $\lambda < 0$ 时， $-\lambda > 0$ ，则说明 ∇f 与 ∇g 同向。这里有个概念要澄清，当我们谈论 ∇g 时，我们是对着函数 $g(x, y)$ 说话，不是对着方程 $g(x, y) = 0$ 说话。方程 $g(x, y) = 0$ 是函数 $g(x, y)$ 用 $z = 0$ 平面截取的等高线。

方程 $g = 0$ 的法向量方向，就是函数 g 的梯度方向 ∇g 。因此，边界极值点 p ，沿着 ∇g 方向移动一点点，函数 g 的值会变大，而且是以最快的速度变大。反之，沿着 $-\nabla g$ 方向移动一点点（注意，只是一点点，理论上就只是一个**无穷小邻域**），函数 g 的值会最快的变小。机器训练中用到的梯度下降法，也是这个原理。

因此，当我们用拉格朗日乘子法，解出边界极值点时 p ，由于 $\nabla \mathcal{L} = 0$ ，保障了点 p 无论向何方移动一点点， \mathcal{L} 的值都不会改变。那么剩下就是个初中生的问题了：

$$L|_p = f + \lambda g = C \quad \rightarrow \quad f = C - \lambda g = C + (-\lambda)g$$

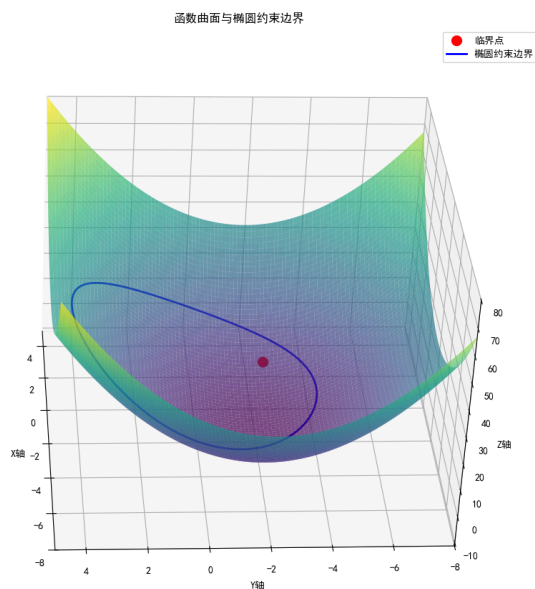
其中 C 是常数， $-\lambda > 0$ 。

现在让边界极值点 p 朝着 $-\nabla g$ 方向移动一点点，也就是向着约束内部移动，由于是梯度反方向，所以函数 g 的值会变小， f 会跟着变小，那么这个边界极值点 p 当然就不是我们要找的全局最小值了。

基于这个原理，我们只保留大于 $\lambda > 0$ 的解，小于0的抛掉。在我们这个例子中，两组解都有 $\lambda < 0$ ，所以全部抛掉。于是只剩下最初找到的内部极值点：

$$p_1 = (1.5, -2.5), \quad f(p_1) = -21.2104$$

下图是本例的可视化效果。后面我们会从几何（微分流形）的角度，彻底搞清楚拉格朗日乘子法。



拉格朗日乘子法，在实际应用中，有一套完整的判定标准，被称作 **KKT条件**：

1. 待优化的问题

KKT条件适用于如下形式的优化问题：

$$\min_x f(x), x \in \mathbb{R}^n \quad \text{s.t.} \quad \begin{cases} g_i(x) \leq 0, & i = 1, \dots, m \\ h_j(x) = 0, & j = 1, \dots, p \end{cases}$$

2. KKT条件的核心内容

1. 原始可行性 (Primal Feasibility)

解必须满足原始约束：

$$g_i(x^*) \leq 0 \quad \forall i, \quad h_j(x^*) = 0 \quad \forall j.$$

2. 对偶可行性 (Dual Feasibility)

不等式约束的拉格朗日乘子必须非负：

$$\lambda_i \geq 0 \quad \forall i.$$

3. 互补松弛性 (Complementary Slackness)

对于每个不等式约束，拉格朗日乘子与约束的乘积为零：

$$\lambda_i g_i(x^*) = 0 \quad \forall i.$$

- 若 $\lambda_i > 0$, 则 $g_i(x^*) = 0$ (约束紧) ;
- 若 $g_i(x^*) < 0$, 则 $\lambda_i = 0$ (约束不起作用) 。

4. 梯度条件 (Stationarity)

目标函数与约束的梯度在最优解处线性相关：

$$\nabla f(x^*) + \sum_{i=1}^m \lambda_i \nabla g_i(x^*) + \sum_{j=1}^p \mu_j \nabla h_j(x^*) = 0.$$

- λ_i 和 μ_j 分别为不等式和等式约束的拉格朗日乘子。

5. 凸性要求 (可选项)

若问题为凸优化 (即 $f(x)$ 凸函数, $g_i(x)$ 凸函数, $h_j(x)$ 仿射函数) , 则KKT条件是**充分必要条件**; 否则仅为必要条件。

用KKT条件解决优化问题，通常步骤如下：

步骤	操作	说明
1	求无约束极值点：解 $\nabla f = 0$, 保留满足所有约束的解	检查是否在可行域内 ($g_i(x) \leq 0$ 且 $h_j(x) = 0$) 。若可行，作为候选解。
2	处理等式约束：解 $\nabla \mathcal{L} = 0$ 和 $h_j(x) = 0$ (忽略不等式约束) , 保留满足 $g_i(x) \leq 0$ 的解	仅当问题无不等式约束时，此步等价于拉格朗日乘子法。
3	对每个不等式约束 $g_i(x) \leq 0$, 分两种情况： a. 松弛 ($\lambda_i = 0$) : 忽略该约束，回到步骤1或2; b. 紧 ($g_i(x) = 0$) : 保留 $\lambda_i \geq 0$, 联立求解 $\nabla \mathcal{L} = 0$ 和 $g_i(x) = 0$ 。	必须穷举所有不等式约束的组合 (如 2^m 种情况, m 为不等式约束数) 。
4	验证所有候选解： a. 检查原始可行性 ($g_i(x) \leq 0$, $h_j(x) = 0$) ; b. 检查对偶可行性 ($\lambda_i \geq 0$) ;	剔除不满足KKT条件的解。

步骤	操作	说明
	c. 互补松弛性 ($\lambda_i g_i(x) = 0$) 。	
5	比较剩余候选解的目标值 $f(x)$ ，选择最小的作为最优解。	若问题非凸，需验证是否为全局最优。

微分流形

虽然**微分流形**这个词听起来可能有些复杂，但实际上，你早就与它打过照面。还记得中学数学课上，老师讲到**数形结合**吗？例如，直线对应于实数集 \mathbb{R} 。但是，当时我们并没有深入理解**实数**的概念，只是知道一些基本的代数运算。随着几何的学习，我们开始理解用坐标表示**点**，进而逐渐意识到：原来直线上的每一个**点**都可以用一个**实数**来描述。

因此，实数集 \mathbb{R} 不仅是一个代数对象，它还有几何性质。将 \mathbb{R} 看作一根笔直的金丝，它的几何特性与我们的直觉一致：它是直的（这是显然的），均匀的（不会忽粗忽细），光滑的（没有尖角或毛刺），而且是稠密的（没有空洞）。因此， \mathbb{R} 被称为**1维欧几里得流形**，它就像测量几何对象的标准尺，我们用它来衡量其他几何对象是否是流形。

如果我们将这根笔直的金丝弯曲，但在局部仍然看起来光滑，那么它就变成了一个非平凡的**1维微分流形**。重要的是：虽然**整体**上它弯曲了，但**局部**上仍然像直线一样。换句话说，流形在局部上和 \mathbb{R} 相似，这正是流形的核心特性。推广到 n **维流形**，就是一个可以在局部和 \mathbb{R}^n 建立一一对应关系的空间。

如何理解“整体弯曲，局部像直线”呢？现代人都知道地球和足球是圆的，但丝毫不影响你认为自家地板是个平面。这个现象，很好的说明了“局部欧几里得”（locally Euclidean）的概念。对于微分流形来说，“局部”是一个抽象概念，指的是在微分结构的尺度（例如 dx ）下，流形的局部邻域看起来与欧几里得空间 \mathbb{R}^n 没有区别。

然而，如果这个金丝被你弄成了尖角或断裂，那它就不再是一个微分流形了。**微分流形要求不仅仅是弯曲，还必须是光滑且可导的**。你熟悉的函数图像，比如：

$$y = \sin x, \quad y = x^2, \quad y = e^x$$

它们的图像在平面上都是光滑的曲线，因此，它们都是1维的微分流形。因此，可以松一口气，虽然“微分流形”这个词听起来很复杂，但其实你从初中开始就在接触它。

描述流形的两种方式

我们不会在此给出流形的严格数学定义，因为它涉及一些抽象的概念和较高的数学知识，感兴趣的读者可以通过专业书籍深入了解。这里只需要记住：流形=集合+结构。如果你了解现代数学中的一些基础概念，你就知道，实数 \mathbb{R} 不仅仅是一个集合，它还携带着丰富的代数结构、拓扑结构和顺序结构等。其中拓扑和微分结构赋予了 \mathbb{R} 几何性质，当我们在平面上可视化它时，它表现为一条直线。

类似地，流形并不只是它在 \mathbb{R}^n 中的图像。你在看到的曲线或曲面只是流形在 \mathbb{R}^n 中的一个“投影”，而流形本身是一个抽象的数学对象，不依赖于任何特定的嵌入空间或可视化方式，它是客观存在的。

这里可能会引发一个哲学性的问题：如果一个抽象概念无法通过任何方式描述或感知，那么它是否仍然是客观存在的？如果从布劳威尔的直觉主义哲学来理解流形，他会更关注流形的“构造性”和“直觉性”。流形的每个部分应该是可构造的，并且在局部上应该能够直观地理解或建构。流形的“存在性”不是通过抽象的公理和定义来证明的，而是通过构造性的展示来理解和确认。因此，流形不仅是一个抽象的几何概念，更是一个可以通过直觉方法和构造步骤加以理解和表达的数学对象。

1. 参数方程

你熟悉的函数

$$y = \sin x$$

其函数图像，就是一个通过参数方程定义的1维微分流形 M ，嵌入 \mathbb{R}^2 中的可视化表象：

$$M = \{(x, \sin x) \in \mathbb{R}^2 \mid x \in \mathbb{R}\}$$

让我们从微分几何的**坐标卡**（coordinate chart）概念来理解这个例子。一个坐标卡记作 (φ, U) ，在本例中这个坐标卡长成这样：

$$\varphi : U \rightarrow \varphi(U) \quad p \mapsto x$$

其中 $U \subset M$ ， $\varphi(U) \subset \mathbb{R}$ ， $p \in M$ ， $x \in \mathbb{R}$ 。

实际上坐标卡就是个函数而已，当然，要求它必须是**微分同胚**的局部映射。所谓微分同胚映射，粗略的讲，就是均有微分结构的两个集合之间的——映射。定义域属于流形 M 的开集，值域属于 \mathbb{R} ，写成经典的函数形式，就是这样：

$$\varphi(p) = x$$

就是将 U 中的点 p 映射到 $\varphi(U)$ 中的点 x 。坐标卡将流形 M 上的点 p ，映射到**参数空间**的点 x ，而参数空间的点，被称作**局部坐标**。

注意，在工程数学中，与纯数学不同，经常会把坐标卡的映射方向反过来，写成：

$$\varphi^{-1} : \mathbb{R} \rightarrow U \quad x \mapsto p$$

由于微分同胚映射是一一映射，因此逆映射也是成立的。之所以工程上喜欢反过来，是因为建模的过程，是通过参数空间的点来构造流形的点，而不是反过来。

除了坐标卡映射（也称局部坐标映射），还有一种关键的映射，称为**嵌入**（embedding）映射，某些简单的 n 维流形，可以嵌入到 $n + 1$ 维欧几里得空间 \mathbb{R}^{n+1} 中。而有个惠特尼（Whitney）定理，证明了任何 n 维流形都可以嵌入到 \mathbb{R}^{2n} 中。

我们能看到函数图像代表的微分流形，都是因为嵌入到 $\mathbb{R}^2, \mathbb{R}^3$ 中，才能够可视化。你想一想，实际上老师在黑板上画出 $y = \sin x$ 的图像时，他就是在作一个嵌入动作。黑板显然就是个二维欧几里得空间，他把 $y = \sin x$ 映射到了 \mathbb{R}^2 中的一个点上，然后把这个点画出来。被流形 M 嵌入的更高维度的欧氏空间 \mathbb{R}^n ，被称作**环境空间**（ambient space）。

如果不嵌入，我们就看不到流形！

在本例中，嵌入映射如下：

$$\iota : M \rightarrow \mathbb{R}^2, \quad p \mapsto (x, \sin x)$$

而参数方程就是：

$$\iota \circ \varphi^{-1} : \mathbb{R} \rightarrow \mathbb{R}^2, \quad x \mapsto (x, \sin x)$$

注意，虽然 $y = \sin x$ 看上去不像是个参数方程形式，但是从微分几何视角，它就是个参数方程。显式的写出来，其实是这样：

$$x = x, \quad y = \sin x$$

由于共用了同一个变量 x ，所以这种写法是不好的，容易导致混淆。更好的写法是：

$$x = t, \quad y = \sin t$$

因为嵌入最低限度也是 \mathbb{R}^2 ，因此参数方程是个**向量值函数**。在咱们例子中有：

$$f(t) = (\iota \circ \varphi^{-1})(t) = (f_1(t), f_2(t)) = (t, \sin t)$$

注意，能够用参数方程表示的微分流形，其实都是些简单的流形。这个简单是针对数学家而言的。工程师遇到的，无论多么复杂的曲面，只要是光滑的，从纯数学的角度看，都是非常简单的流形。

最简单的流形，就是像 $y = f(x)$ 这种，不用分段，用一个连续的函数就能表示，因而用一个参数方程就能将其嵌入到 \mathbb{R}^2 中。但是微分流形的核心思想恰好是反过来的，它关注的其实是分段函数，甚至到无穷多段。其次，它每一段函数管的范围可以小到一段微分的范围。这一小段范围，称之为**局部**。

这两个特性，通过坐标卡 (φ, U) 的概念体现。也就是说，典型的微分流形，能够做到：由很多个乃至无穷多坐标卡 (φ_i, U_i) 组成，每个坐标卡中的函数 φ_i ，只管自己的无穷小的一亩三分地 U_i 。并且，每个坐标卡，必然与其他至少一个坐标卡有非空交集。

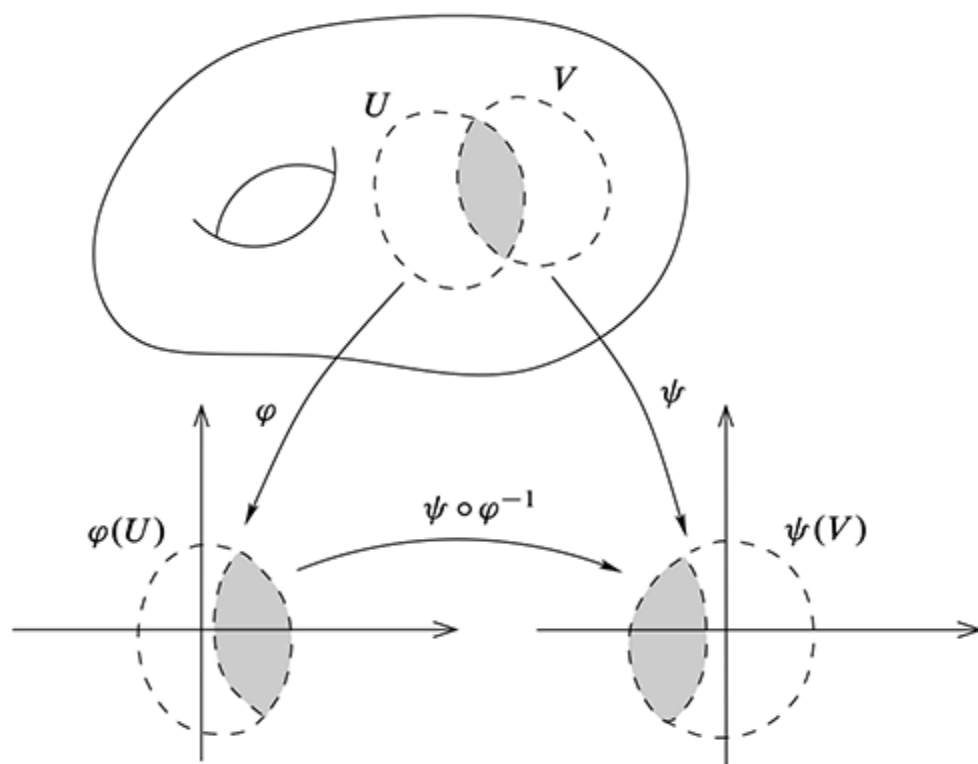
坐标卡的一些特性：

1. 通常来说， U 只是 M 的某个**开集**，不能覆盖整个 M 。 φ 是一个**局部微分同胚映射**，这正是流形的局部欧几里得性质的严格表述。由于一个 φ 不能覆盖整个 M ，所以我们需要很多个坐标卡 (φ_i, U_i) 来覆盖 M 中的所有点，这些坐标卡的集合叫做**图册** (atlas)。
2. 接下来很关键，是微分几何能够变魔法的根本原因：坐标卡之间，必须要**相容** (compatible)。如果两个坐标卡 $(\varphi, U), (\psi, V)$ 满足条件

$$U \cap V \neq \emptyset$$

则在交集 $U \cap V$ 上，两个坐标卡之间的过渡函数 $\psi \circ \varphi^{-1}$ 必须是丝滑 (smooth) 的，即它是 \mathbb{R}^n 到 \mathbb{R}^n 之间的光滑映射：

$$\psi \circ \varphi^{-1} : \varphi(U \cap V) \rightarrow \psi(U \cap V)$$



看图说话，就很容易理解了。看着很亲切，仿佛回到了幼儿园时代，或者布尔巴基小学。注意看， $\psi \circ \varphi^{-1}$ 其实是 \mathbb{R}^n 到 \mathbb{R}^n 的映射。这正是微分几何的小把戏的关键之处，意味着通过过渡函数，我们可以把流形从一个局部坐标系统平滑地转换到另一个坐标系统。

这种搞法，使得我们熟悉的欧氏空间的连续可导等性质，代替了流形 M 的性质。也就是说，由于规定流形必须满足局部欧几里得性质，而我们又用坐标卡，每次只描述流形的一个局部子集，所以我们可以用欧氏空间的微积分理论来描述流形的性质。整个事情就这么简单，没有什么神奇的地方。

流形的定义不依赖嵌入（如广义相对论的四维时空流形无环境空间）。内蕴几何通过切空间、黎曼度量等工具描述流形性质，无需嵌入。

但是在工程实践中，如果一个流形能够嵌入 R^n 中，我们总是很乐意这么干。嵌入后很方便研究整体性质，不干是傻子。而我们面对的凸优化问题，也是在 R^n 中求解的，所以我们需要将流形嵌入到 R^n 中，才能用凸优化方法求解。所以，不用担心黎曼几何的内蕴性质，我们并不涉及这些内容。

对初学者，可能有个地方会产生疑惑，就是流形上的点 p ，到底是怎么回事，怎么用数值来表示？实际上这是个误区， p 是个抽象的概念，不能直接观察它。我们要么前推，用坐标卡将 p 映射到参数空间。要么后推，用嵌入将 p 映射到环境空间。而实际上你见到的一连串眼花缭乱

的数学操作，都是通过复合映射：

$$\iota \circ \varphi^{-1} : \mathbb{R} \rightarrow M \rightarrow \mathbb{R}^2, \quad x \mapsto p \mapsto (x, \sin x)$$

来运作的，中间的p是个黑箱，并无实际数据对应它，只是个概念。所以你真正看到的数学表达式，仅仅是：

$$\iota \circ \varphi^{-1} : \mathbb{R} \rightarrow \mathbb{R}^2, \quad t \mapsto (x = t, y = \sin t)$$

这正是一开始入门微分几何的难点。正所谓：道可道，非常道；名可名，非常名。无名天地之始，有名万物之母。故常无欲，以观其妙；常有欲，以观其徼。此两者同出而异名，同谓之玄，玄之又玄，众妙之门

空间	符号	点	表现形式
参数空间	\mathbb{R}^n	(t_1, t_2, \cdots)	局部坐标
流形	M	p	抽象的概念，无法直接观察
环境空间	$\mathbb{R}^{n+1} - \mathbb{R}^{2n}$	(x, y, z, \cdots)	环境坐标

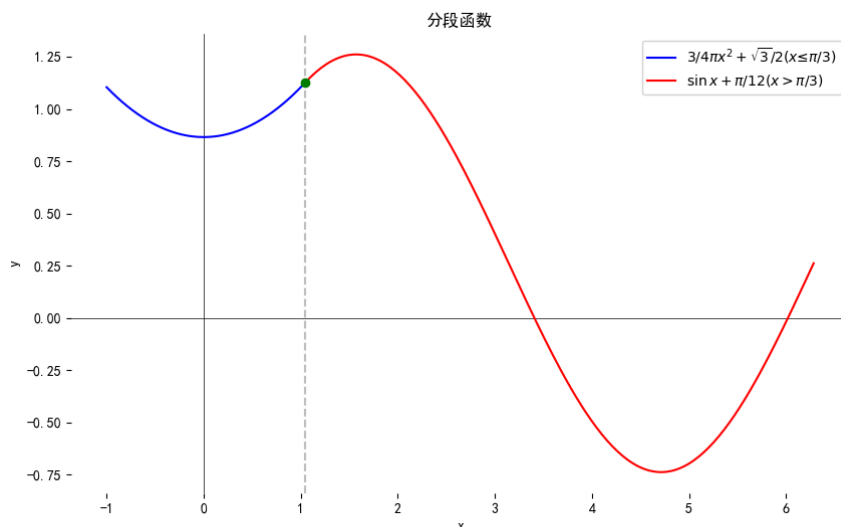
流形上的点p无法直接观察！

分段构造例子

让我们来看一个，对于中学生很头疼的分段函数的例子：

$$y = \begin{cases} 3/4\pi x^2 + \sqrt{3}/2 & x \leq \pi/3 \\ \sin x + \pi/12 & \pi/3 < x \end{cases}$$

下面是函数的图像：



我给出了两个坐标卡 $(\varphi_1, U_1), (\varphi_2, U_2)$, 其中:

$$\varphi_1(U_1) = (-\infty, \pi/3), \quad \varphi_2(U_2) = (\pi/3, \infty)$$

注意, 坐标卡要求定义域为开集, 因此局部坐标 $\pi/3$ 对应的那一个点, 并没有被覆盖到。而且很明显, $\varphi_1(U_1) \cap \varphi_2(U_2) = \emptyset$, 根据微分同胚映射的定义, 可知 $U \cap V = \emptyset$, 因此这两个坐标卡并无交集。因此需要构造第三个坐标卡 (φ_3, U_3) :

$$\varphi_3(U_3) = (\pi/3 - \epsilon, \pi/3 + \epsilon)$$

其中 $\epsilon > 0$ 为小正数; 这样的话, 整个流形的图册就由3个坐标卡组成:

$$\text{atlas} = \{(\varphi_1, U_1), (\varphi_2, U_2), (\varphi_3, U_3)\}$$

可以验证一下是否微分流形, 先看最基本的是否连续:

$$\left. \frac{3}{4\pi}x^2 + \frac{\sqrt{3}}{2} \right|_{x=\pi/3} = \frac{\pi}{12} + \frac{\sqrt{3}}{2}, \quad \left. \sin x + \frac{\pi}{12} \right|_{x=\pi/3} = \frac{\sqrt{3}}{2} + \frac{\pi}{12}$$

再验证 $x = \pi/3$ 处左右导数是否相等:

$$\left. \frac{d}{dx} \left(\frac{3}{4\pi}x^2 + \frac{\sqrt{3}}{2} \right) \right|_{x=\pi/3} = \frac{3}{2\pi} \cdot \frac{\pi}{3} = \frac{1}{2}$$

$$\left. \frac{d}{dx} \left(\sin x + \frac{\pi}{12} \right) \right|_{x=\pi/3} = \cos \left(\frac{\pi}{3} \right) = \frac{1}{2}$$

这表明转移映射在邻域内一阶光滑。再验证二阶导数：

$$\begin{aligned} \left. \frac{d^2}{dx^2} \left(\frac{3}{4\pi} x^2 + \frac{\sqrt{3}}{2} \right) \right|_{x=\pi/3} &= \frac{3}{2\pi} \\ \left. \frac{d^2}{dx^2} \left(\sin x + \frac{\pi}{12} \right) \right|_{x=\pi/3} &= -\sin \left(\frac{\pi}{3} \right) = -\frac{\sqrt{3}}{2} \end{aligned}$$

显然二阶导数不相等。结论是：我们构造的流形在 $x = \pi/3$ 处，一阶导数存在且连续，这种微分流形叫做** C^1 流形**，而不是光滑流形，那个要求无穷阶可导，记作 C^∞ 流形。

2维流形例子

用 $y = \sin x$ 举例，过于平凡，会忽略很多细节。接下来我们看一下典型的2维流形，就是单位球面，记作 S^2 ，如何用参数方程描述：

$$S^2 = \{ \sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta \} \quad \theta \in [0, \pi], \phi \in [0, 2\pi]$$

这是工程上的典型表达方式，如果严格按照微分几何的符号规范，需要改成：

$$x^1 = \theta, x^2 = \phi$$

于是规范的写法是这样：

$$S^2 = \{ \sin x^1 \cos x^2, \sin x^1 \sin x^2, \cos x^1 \} \quad x^1 \in [0, \pi), x^2 \in [0, 2\pi)$$

微分几何的惯例，局部坐标的索引用上标。一开始可能不习惯 x^2 这种搞法，但习惯了之后，就会觉得很自然。用上标索引有两个好处：一是索引上下位置表明了向量的类型（逆变/协变），二是用索引可以采用爱因斯坦约定简化公式。

于是参数空间的局部坐标，就可以写成向量形式：

$$\boldsymbol{x} = \left(x^1 \frac{\partial}{\partial x^1}, x^2 \frac{\partial}{\partial x^2} \right) = (x^1 \partial_1, x^2 \partial_2), \quad \boldsymbol{x} \in \mathbb{R}^2$$

另外，当我们说流形 M 与 \mathbb{R}^2 微分同胚时，其实更准确的意思是说， M 与 仿射空间 \mathbb{R}^2 同胚。一般情况下，我们不区分仿射空间 \mathbb{A}^n 与 线性空间 \mathbb{R}^n ，而是统称为欧氏空间 \mathbb{R}^n 。因此，向量 \boldsymbol{x} 也可以看作仿射空间的一个点 x 。此时写作：

$$x = x^1 \frac{\partial}{\partial x^1} + x^2 \frac{\partial}{\partial x^2} = x^1 \partial_1 + x^2 \partial_2 = x^\mu \partial_\mu \quad x \in \mathbb{A}^2$$

其中：

$$x^\mu \partial_\mu = \sum_{\mu=1}^2 x^\mu \partial_\mu = x^1 \partial_1 + x^2 \partial_2$$

用到了爱因斯坦约定。这样一写，是否立刻显得高端大气上档次呢？因为用了希腊字母，看起来很像广义相对论理论中的符号系统。的确是的，这里用希腊字母表示**分量指标**，是彭罗斯发明的张量**抽象指标**表示法中的规范。什么时候用拉丁字母，什么时候用希腊字母，都是有讲究的。后面我们会详细谈到这些内容。

另外：

$$\frac{\partial}{\partial x^1} = \partial_1 = \mathbf{e}_1, \quad \frac{\partial}{\partial x^2} = \partial_2 = \mathbf{e}_2$$

这也是微分几何中，表达**逆变基**（这些概念后面会解释）的规范写法。于是也可以写成：

$$\boldsymbol{x} = x^1 \mathbf{e}_1 + x^2 \mathbf{e}_2 = x^\mu \mathbf{e}_\mu \quad \mu = 1, 2$$

这个球面 S^2 ，通过参数方程，被嵌入到 \mathbb{R}^3 后，其在环境空间的坐标，就可以写成：

$$\begin{aligned}y^1 &= y^1(x^1, x^2) = \sin x^1 \cos x^2 \\y^2 &= y^2(x^1, x^2) = \sin x^1 \sin x^2 \\y^3 &= y^3(x^1, x^2) = \cos x^1\end{aligned}$$

这样就可以很方便的改写成向量形式：

$$\mathbf{y}(\mathbf{x}) = (y^1(\mathbf{x}), y^2(\mathbf{x}), y^3(\mathbf{x}))$$

注意看， $\mathbf{x} \in \mathbb{R}^2$ ， $\mathbf{y}(\mathbf{x}) \in \mathbb{R}^3$ 。这就是经典的**向量微积分**，如果你不接触流形理论，照样也可以算题，也用雅可比矩阵：

$$J = \begin{bmatrix} \frac{\partial y^1}{\partial x^1} & \frac{\partial y^1}{\partial x^2} \\ \frac{\partial y^2}{\partial x^1} & \frac{\partial y^2}{\partial x^2} \\ \frac{\partial y^3}{\partial x^1} & \frac{\partial y^3}{\partial x^2} \end{bmatrix}$$

进行坐标变换。但实际上，映射：

$$\mathbf{f} : \mathbb{R}^2 \rightarrow \mathbb{R}^3, \quad \mathbf{x} \mapsto \mathbf{y}$$

中间藏着一个黑箱的 S^2 流形：

$$\mathbf{f} : \mathbb{R}^2 \rightarrow S^2 \rightarrow \mathbb{R}^3, \quad \mathbf{x} \mapsto p \mapsto \mathbf{y}$$

而这个点 p ，我们是看不到的。作为工程师，也不会去关心它。到这里，应该回到了你比较熟悉的符号系统了，是不是忽然觉得自己又行了，又会了。诺伊曼说过：

Young man, in mathematics you don't understand things. You just get used to them.
——John von Neumann

这句话的意思是：在数学中，有些东西一开始你是无法真正“理解”的（至少不是用日常直觉理解），你只能通过不断接触、操作、运算，最终对它**习惯了，感觉自然了**，从而形成了你的“理解”。

2. 隐式方程

隐式方程是一种描述流形的方式，还是以 S^2 为例：

$$x^2 + y^2 + z^2 = 1$$

它实际上是一个代数曲面的零点集：

$$h(x, y, z) = x^2 + y^2 + z^2 - 1 = 0$$

在机器学习中，隐式方程通常用于描述**约束**（constraints）。而从微分几何的角度来看， n 元隐式方程描述了一个 $n - 1$ 维流形被嵌入 R^n 中。如果我们写成：

$$h(\boldsymbol{x}) = 0, \quad \boldsymbol{x} = (x^1, x^2, x^3) \in \mathbb{R}^3$$

看得很清楚，三元隐式方程描述了一个二维流形 S^2 被嵌入到 \mathbb{R}^3 中。因此，我们必须要有微分流形的基础知识，才能更好的理解凸优化理论。

3. 两种方式对比

名称	坐标	方式	数学表述	适用场景
参数方程	局部	定义	$\phi : \mathbb{R}^k \rightarrow \mathbb{R}^n, \quad \boldsymbol{u} \mapsto \boldsymbol{x} = \phi(\boldsymbol{u})$	精确控制自由度或避免奇点
隐式方程	全局	约束	$M = \{\boldsymbol{x} \in \mathbb{R}^n \mid h(\boldsymbol{x}) = 0\}$	快速原型设计或复杂约束

流形的基本性质

1. 切空间：

流形上任一点 p 的切空间记作 $T_p M$ ，从这个符号就可以看出，它是流形的局部性质，每个点都有自己的切空间。有以下几种方式定义 $T_p M$ ：

- 几何定义：

设 $\gamma : (-\epsilon, \epsilon) \rightarrow M$ 是满足 $\gamma(0) = p$ 的光滑曲线。两条曲线 γ_1, γ_2 在 p 点**等价**（记为 $\gamma_1 \sim \gamma_2$ ），若它们在某个（进而所有）坐标图 (U, φ) 下满足：

$$\left. \frac{d}{dt}(\varphi \circ \gamma_1(t)) \right|_{t=0} = \left. \frac{d}{dt}(\varphi \circ \gamma_2(t)) \right|_{t=0}$$

切空间 $T_p M$ 是这些等价类的集合，每个等价类 $[\gamma]$ 称为一个切向量。

- **坐标表示** 在坐标卡 (φ, U) 下，若 $\varphi = (x^1, \dots, x^n)$ ，则切空间有基底：

$$\left. \frac{\partial}{\partial x^i} \right|_p \quad (i = 1, \dots, n)$$

其中 $\left. \frac{\partial}{\partial x^i} \right|_p$ 是沿坐标曲线 x^i 的方向导数。任意切向量可表示为：

$$v = \sum_{i=1}^n v^i \left. \frac{\partial}{\partial x^i} \right|_p$$

- **代数导数**：切向量是满足莱布尼茨律的线性算子 $v : C^\infty(M) \rightarrow \mathbb{R}$ ：

$$v(fg) \Big|_p = v(f) \Big|_p g(p) + f(p) v(g) \Big|_p \quad \forall f, g \in C^\infty(M).$$

所有这样的算子构成 $T_p M$ 。

余切空间：

余切空间 $T_p^* M$ 定义为切空间的^{*}对偶空间，即所有线性泛函 ω 的集合：

$$T_p^* M = \{\omega : T_p M \rightarrow \mathbb{R}\}$$

- **微分形式定义**：

对任意 $f \in C^\infty(M)$ ，其微分 $df \Big|_p$ 是一个余切向量，定义为：

$$df \Big|_p(v) = v(f) \Big|_p, \quad \forall v \in T_p M.$$

所有这样的微分生成 $T_p^* M$ 。

- **坐标表示**

在坐标图 (U, φ) 下，余切空间有对偶基底 $dx^i \Big|_p$ ，满足：

$$dx^i \Big|_p \left(\frac{\partial}{\partial x^j} \Big|_p \right) = \delta_j^i.$$

任意余切向量可表示为：

$$\omega = \sum_{i=1}^n \omega_i dx^i \Big|_p.$$

法空间

在微分几何中，给定一个**嵌入子流形 (Embedded Submanifold)** M 包含于一个更大的流形如 \mathbb{R}^n ，法空间 $N_p M$ 的定义如下：

$$N_p M := \{v \in T_p \mathbb{R}^n \mid \forall w \in T_p M, \langle v, w \rangle_g = 0\}$$

法空间的定义依赖于子流形 M 嵌入一个更大的环境空间 \mathbb{R}^n ， M 的法空间是切空间在 \mathbb{R}^n 的切空间中的正交补，即：

$$N_p M \oplus T_p M = T_p \mathbb{R}^n$$

由于 \mathbb{R}^n 的切空间与自身同构（甚至可以认为就相等）：

$$T_p \mathbb{R}^n \cong \mathbb{R}^n \quad \text{or} \quad T_p \mathbb{R}^n = \mathbb{R}^n$$

因此，理解了之后，对于流形 M 嵌入环境空间 \mathbb{R}^n (符合绝大多实际的物理现象，经济学模型，以及机器学习模型等，对工程师来说足够了)的情况, 也可以简写为：

$$N_p M \oplus T_p M = \mathbb{R}^n$$

这一堆定义有点抽象，那么我们就用 S^2 的例子，来尝试理解切空间。前面讲过，流形上的点 p 是无法直接看到或触摸到的。如果我们想要表达它，就必须通过映射将其与某个参数空间的局部坐标或嵌入空间中的环境坐标关联起来。

在这里，术语使用上可能会让人感到困惑。部分文献中将环境坐标称作“全局坐标”，而在主流微分几何文献中，通常将参数空间中的坐标称为“局部坐标”。事实上，环境坐标在某种意义上也是

局部坐标。因此，有时会产生困惑，不知道该如何区分。我们在此规定如下：

- **参数坐标**：通常指流形的参数化所使用的坐标，也叫**局部坐标**。这些坐标描述的是流形的某个局部区域。
- **环境坐标**：是指流形嵌入到某个更高维的空间（比如 \mathbb{R}^n ）后的坐标。虽然有些文献将其称为“全局坐标”，但从严格意义上讲，它本质上也是局部坐标，因为它通过嵌入映射与环境空间的坐标系关联的。

以隐式方程为例：在某些情况下，流形只有环境坐标，没有明确的参数坐标。对于单位球面 S^2 ，我们用环境坐标 $\mathbf{x} = (x^1, x^2, x^3)$ 来表示其点 p 。这里的 \mathbf{x} 只是流形上点 p 的一种表示方式。需要反复强调的是，环境坐标并不等同于流形的实际点，它仅仅是流形点的一种表示方式。

这种表示方式就像我们在观看电影《哪吒闹海》时，看到屏幕上哪吒的动画形象。你可以指着屏幕上的哪吒说：“这是哪吒”。虽然这句话没错，但我们知道，哪吒是一个艺术形象，而动画只是通过一种可视化的方式展示了这个形象。对于流形上的点，环境坐标和参数坐标是不同的表示方式，但它们都只是通过某种映射展示了流形上的点。

理解了这一点之后，我们就不再纠结形式主义，就直接用环境坐标 \mathbf{x} ，表示流形上的点 p 。我们随便选个点：

$$x^1 = x = 0.3827, \quad x^2 = y = -0.8359$$

$x^3 = z$ 要通过隐式方程：

$$h(\mathbf{x}) = x^2 + y^2 + z^2 - 1 = 0$$

算出来。这里其实挺尴尬的，你会发现在实际运算中，如果采用上标符号，会令人非常不适：

$$x^3 = z = \sqrt{1 - 0.3827^2 - (-0.8359)^2} = 0.3935$$

这也说明了工业界有自己的符号系统，术语标准，与学术界不同，有其自己的理由。我们这里用切空间的第二种定义，坐标表示的方式来计算。

$$\mathbf{x} = (x^1, x^2, x^3) = (0.3827, -0.8359, 0.3935)$$

隐函数 $f(\mathbf{x})$ 在 p 点的全微分为：

$$\begin{aligned}
df|_p &= \left(\frac{\partial h}{\partial x^1} dx^1 + \frac{\partial h}{\partial x^2} dx^2 + \frac{\partial h}{\partial x^3} dx^3 \right)_p \\
&= (2x^1 dx^1 + 2x^2 dx^2 + 2x^3 dx^3)_p \\
&= 0.7654 dx^1 - 1.6718 dx^2 + 0.7869 dx^3
\end{aligned}$$

由于隐函数总是等于0（常函数），那么其全微分在整个定义域都恒等于0：

$$df \equiv 0$$

在我们任意的点 p 也应不例外。可是我们算出来的结果，是这样：

$$df|_p = 0.7654 dx^1 - 1.6718 dx^2 + 0.7869 dx^3$$

这玩意怎么等于零呢？你想明白了吗？不妨回顾一下我们最开始学微积分的内容，一个单变量函数，并且是常函数：

$$y = f(x) = C$$

画在平面直角坐标系中，就是根水平直线，是个平凡的函数。若 $C = 0$ ，那更是平凡的不能再平凡了。那么有：

$$df \equiv 0$$

你有想过，这个等式，到底表达了什么几何意义呢？注意直线本身也是流形，也就是说，任何一个单变量函数，都是定义在流形 \mathbb{R} 上的，即 x 轴是个流形。因此 $df \equiv 0$ 是不是说，在流形 \mathbb{R} 上任一点 p 处，如果 p 在 x 轴上移动了一个微分的距离 dx ，相应的函数值并不会变化，即 $df = 0$ 在任意点 p 处都成立。如果不是常函数，例如：

$$y = f(x) = ax^2 - bx + c, \quad a > 0$$

那么显然有：

$$df = (2ax - b)dx$$

要想 $df = 0$, 必须要 $2ax - b = 0$ 。因此得到:

$$x = \frac{b}{2a}$$

这是抛物线的最低点 $p(b/2a, -b^2/2a + c)$ 。这就要求必须在流形 \mathbb{R} 上的点 p 即 $(x = b/2a)$ 处, 沿着 x 轴移动一个微分的距离 dx , 函数值并不会改变, 即:

$$df|_p = 0$$

比较一下常函数 $f(x) = C$ 与 抛物线函数 $f(x) = ax^2 - bx + c$, 它们的微分 df , 共同点是什么? 区别又是什么? 共同点就是: 当底流形 (下轴) 上某点 p 移动一个微分的距离 dx 范围内, 函数值 $f(x)$ 并不会变化。

认识到这一点之后, 现在我们回过头看一下, 隐函数 $f(\mathbf{x}) = 0$ 意味着什么? 它也是个常函数, 它的底流形 (自变量的定义域) 就是 S^2 , 这被称作**流形上的微积分**。虽然 $h(\mathbf{x})$ 的自变量看上去有3个分量, 但是自由度只有2个, 3个分量并不独立, x^1, x^2, x^3 并不能全部自由自在的在 R^3 随意变换。我们刚才随意找到 p 点过程, 其实就清楚的显示了这一点。先是任意找了 $x^1 \in [-1, 1], x^2 \in [-1, 1]$, 然后 x^3 是通过约束方程找到的。因为 $\mathbf{x} \in S^2$, 它必须在球面上移动。

接下来的问题就是, 怎么定义点 p 在 S^2 上移动了一个微分的距离, 从 \mathbf{x} 移到了 $\mathbf{x} + d\mathbf{x}$ 。正如微积分的核心思想是化曲为直, 微分流形也一样。只要谈到微分, 由于微分流形要求局部与 \mathbb{R}^n 微分同胚, 所谓局部欧几里得性, 指的就是, 即使流形的表面是弯曲的, 但是在一个微分的范围内, 可以安全的认为就是直的。

于是几何学家干脆也不装了, 它们就直接定义了一个在 p 点与流形相切的欧几里得空间 \mathbb{R}^n , 维度 n 与流形的维度相同, 称之为切空间, 记作 $T_p M$ 。在本例中, 由于 $\dim(S^2) = 2$, 因此有:

$$T_p S^2 = \mathbb{R}^2$$

于是就将点 p 在 S^2 的移动, 变成了在 $T_p S^2$ 平面内移动。有人会困惑, 这样一来, 不就脱离了 S^2 表面了吗? 其实个中奥秘, 我也不懂, 总之就是不掉链子。假设你有一个拱形的巨型建筑表面, 想美化一下。贴马赛克, 每块马赛克都是很小的正方形, 你一定能贴的看上去很完美, 完全贴合这个建筑的拱形表面。这就是局部欧几里得性质的威力。大概的原因是因为, 随着点 p 的移动, 切空间 $T_p M$ 也跟着变。想象一下, 有个光滑曲线, 任选一点 p , 做一条切线。然后随

着点 p 的移动，切线也在不断的变化。你自己脑补一下，切空间就是这样，随着点 p 的移动，切空间也在不断的变化。

假设你现在对切空间 $T_p M$ 的概念有了初步认识，自然就会想着怎么表达它？Talk is cheap, show me the equation. 我们刚才随意选了 S^2 上一个点 p ，算出来一组数。一般性的写成：

$$\omega = \omega_1 dx^1 + \omega_2 dx^2 + \omega_3 dx^3 = \omega_\mu dx^\mu$$

也即是说：

$$df = \omega = \omega_\mu dx^\mu$$

我们又知道应该有：

$$df = 0$$

但是我们算出来的东东，看上去怎么也不像是0的样子。奥妙在哪呢？考虑 \mathbb{R}^3 中，过一个点 p 的点法式平面方程：

$$A(x - x_p) + B(y - y_p) + C(z - z_p) = 0$$

有人可能认为，点法式平面方程，可以写成向量点积的形式：

$$(A, B, C) \cdot (x - x_p, y - y_p, z - z_p) = 0$$

也不能说你错，如果没有学习微分几何，这么干是可以的。但是学了微分几何，再这样就略显不太专业，后面我们会详细解释其中的端倪。

如果我们知道系数 A, B, C ，实际上这个平面就已经确定了，对吧？而实际上前面，我们的确算出来了 A, B, C ，它们就是 $\omega_1, \omega_2, \omega_3$ 。即：

$$(A, B, C) = (\omega_1, \omega_2, \omega_3)_p = \left(\frac{\partial h}{\partial x^1}, \frac{\partial h}{\partial x^2}, \frac{\partial h}{\partial x^3} \right)_p$$

而：

$$X = (x - x_p, y - y_p, z - z_p) \in T_p S^2$$

你现在看明白了吧？点法式平面方程，就可以写成：

$$df(X) = 0$$

因此，对于流形上的光滑函数 f ，物理学一般称其为光滑标量场，如果是个常函数 $f(\mathbf{x}) = C$ ，则 $df|_p = 0$ 的意思是，在 p 点处，沿着 $T_p S^2$ 平面内移动，函数值 $f(\mathbf{x})$ 不会变化。实际上这正是导数的定义，并没有什么神秘的地方。只不过，定义在平凡流形 R^n 上的光滑函数的导数，由于其平直性，导致：

$$T_p \mathbb{R}^n = \mathbb{R}^n$$

并且其局部坐标覆盖全局，只有一个平凡的恒等映射坐标卡，因此切向量 X 所在的 $T_p \mathbb{R}^n$ 是已知的。例如，对于定义在 \mathbb{R}^2 上的光滑函数 $f(x, y)$ ，任一点 $p(x_0, y_0)$ ，其切向量 X 可以表示为：

$$X = (X_1, X_2, 0)$$

对应的法向量永远垂直于xy坐标平面，而切空间 $T_p \mathbb{R}^2$ 就是xy坐标平面本身。因而这是平凡的，不需要显式的写出来。故简写为： $df = 0$ 。不妨还是用单位球的上半部分来说说事：

$$f(x, y) = 1 - \sqrt{x^2 + y^2}$$

记住，学了微分几何之后，永远从流形的角度，来考虑函数。这个函数，其实是定义在平凡流形 \mathbb{R}^2 上的光滑函数。不要与前面谈论的参数化方程定义流形的事情搞混了。事情是这样的，好比俄罗斯套娃，在底流形（定义域）上定义光滑函数，则函数图像会构造出一个新的流形。就是说：

$$f : M \rightarrow \mathbb{R} \quad p \mapsto f(p)$$

如果 f 是个光滑函数，M是个n维流形，那么：

$$\Gamma_f = \{(p, f(p)) \in R^{n+1} \mid p \in M, f(p) \in \mathbb{R}\}$$

则 Γ_f 是个嵌入到 \mathbb{R}^{n+1} 中的 n 维流形。因此，对于任意一个光滑函数 $f(\mathbf{x})$ ，要看其定义域。如果定义域是 \mathbb{R}^n 的子集，即 $\mathbf{x} \in \mathbb{R}^n$ ，则因为 \mathbb{R}^n 是个平凡流形，这就是经典微积分。如果定义域属于一个非平凡的流形 M ，即 $\mathbf{x} \in M$ ，则称 $f(\mathbf{x})$ 是定义在 M 上的光滑函数。由于底流形（定义域）的非平凡性，此时就需要用到微分几何的概念了。

好的，现在有一个定义在 \mathbb{R}^2 上的光滑函数 $f(x, y)$ ：

$$f(x, y) = 1 - \sqrt{x^2 + y^2}, \quad (x, y) \in \mathbb{R}^2$$

让我们来看一下 $df = 0$ 到底意味着什么：

$$df = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy = -\frac{x}{\sqrt{x^2 + y^2}} dx - \frac{y}{\sqrt{x^2 + y^2}} dy$$

是否就意味着：

$$-\frac{x}{\sqrt{x^2 + y^2}} = 0, \quad -\frac{y}{\sqrt{x^2 + y^2}} = 0$$

也就意味着点法式平面方程的系数为：

$$A = 0, \quad B = 0, \quad C \neq 0$$

这正是半球面的北极点，其法线指向 z 轴正方向，只有这一点才能满足 $df = 0$ 。怎么用微分几何的观点，解释这件事情呢？其实 $df = 0$ 就是 $df(X) = 0$ 的意思。由于 \mathbb{R}^n 的平凡性，省略了后面的 X 而已。这个现象，就相当于 $x = 1x$ ，但我们不会写前面平凡的系数1，因为它是多余的。在这个例子中，由于底流形（ xy 平面）上的任何点，被限制在2维子空间，只能写成：

$$X = (x, y, 0)$$

则

$$\begin{aligned}
df(X) &= \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, C \neq 0 \right) \cdot (x, y, 0) \\
&= -\frac{x^2}{\sqrt{x^2 + y^2}} - \frac{y^2}{\sqrt{x^2 + y^2}} \\
&= -\sqrt{x^2 + y^2} \\
&= 0
\end{aligned}$$

即：

$$df(X) = 0 \iff x^2 + y^2 = 0 \iff x = 0, y = 0$$

这和我们以前学的知识兼容，半球面的北极点，其法线指向z轴正方向，只有这一点才能满足 $df = 0$ 。

到了这里，假设读者对切空间 $T_p M$ 的概念有了初步的认识，我们已经知道，在流形 M 嵌入 \mathbb{R}^n 的前提下，通过全微分 $df = 0$ ，可以得到流形的法向量，进而通过点法式关系，确定切空间 $T_p M$ 。不过，还是感觉有点隔靴搔痒，我们希望构造出一组正交基，来表示切空间 $T_p M$ 。

具体方法为，任选一个向量（我们混用点和向量，在仿射空间这是合法的。点等价于从原点到这个点的向量） $\mathbf{a} = (a_1, a_2, a_3)$ ，并且暂时将 $df = \omega_\mu dx^\mu$ 中的 ω_μ 也看作向量（以后会明白它不是一般意义上的向量，但是从更高的观点看，它又是向量。或许这就是，认知提升过程中的否定之否定吧）。于是 $\boldsymbol{\omega} = (\omega_1, \omega_2, \omega_3)$ ，要求 \mathbf{a} 不能与 $\boldsymbol{\omega}$ 线性相关（平行），即：

$$\mathbf{a} \neq \lambda \boldsymbol{\omega}$$

于是 \mathbf{a} 可以做矩形分解，其中一个分量与 $\boldsymbol{\omega}$ 平行，另一个分量与 $\boldsymbol{\omega}$ 正交。其中正交分量 \mathbf{v} 就是我们要找的切向量。具体计算公式如下：

$$\mathbf{a} = \frac{\mathbf{a} \cdot \boldsymbol{\omega}}{\|\boldsymbol{\omega}\|^2} \boldsymbol{\omega} + \mathbf{v}$$

看上去有点复杂，如果我们先将 $\boldsymbol{\omega}$ 归一化，就会简单很多：

$$\tilde{\boldsymbol{\omega}} = \frac{\boldsymbol{\omega}}{\|\boldsymbol{\omega}\|}$$

则有：

$$\mathbf{a} = (\mathbf{a} \cdot \tilde{\boldsymbol{\omega}}) \tilde{\boldsymbol{\omega}} + \mathbf{v}$$

我们随便选一个向量 $\mathbf{a} = (1, 0, 0)$ ，加上之前得到的 $\boldsymbol{\omega} = (0.7654, -1.6718, 0.7869)$ ，就可以计算了。注意，这里只写了分量，没有带基（基是一个很大的话题，后面会详述）。

$$\begin{aligned}\mathbf{v} &= \mathbf{a} - (\mathbf{a} \cdot \tilde{\boldsymbol{\omega}}) \tilde{\boldsymbol{\omega}} \\ &= (1, 0, 0) - \frac{(1, 0, 0) \cdot (0.7654, -1.6718, 0.7869)}{0.7654^2 + (-1.6718)^2 + 0.7869^2} (0.7654, -1.6718, 0.7869) \\ &= (0.8535, 0.3199, -0.1506)\end{aligned}$$

检验一下是否正交：

$$\boldsymbol{\omega} \cdot \mathbf{v} = 3.36\text{e-}18 \approx 0$$

验证了 \mathbf{v} 确实是切向量。接下来用叉积构造第二个切向量 \mathbf{u} ，这会使得 $\mathbf{u}, \mathbf{v}, \boldsymbol{\omega}$ 三者两两正交：

$$\begin{aligned}\mathbf{u} &= \boldsymbol{\omega} \times \mathbf{v} \\ &= (0.7654, -1.6718, 0.7869) \times (0.8535, 0.3199, -0.1506) \\ &= (0, 0.7869, 1.6718)\end{aligned}$$

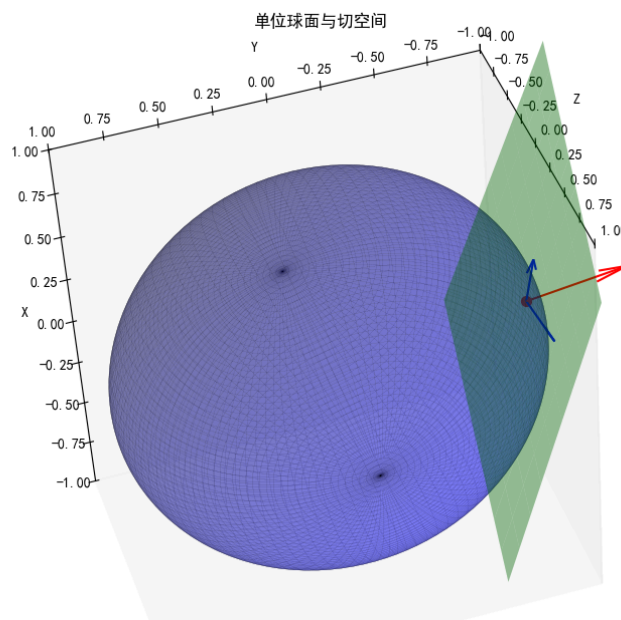
验算一下：

$$\boldsymbol{\omega} \cdot \mathbf{u} = -8.9740\text{e-}17 \approx 0, \quad \mathbf{u} \cdot \mathbf{v} = 2.0937\text{e-}17 \approx 0, \quad \mathbf{v} \cdot \boldsymbol{\omega} = 3.3576\text{e-}18 \approx 0$$

其中 $\mathbf{u}, \mathbf{v} \in T_p S^2$ ，或者说，两个线性无关的向量 \mathbf{u}, \mathbf{v} 张成了切空间 $T_p S^2$ 。实际上，我们可以做的更完美，将三个向量归一化后：

$$\begin{aligned}\tilde{\boldsymbol{\omega}} &= (0.3827, -0.8359, 0.3935) \\ \tilde{\mathbf{u}} &= (0, 0.4259, 0.9048) \\ \tilde{\mathbf{v}} &= (0.9239, 0.3463, -0.1630)\end{aligned}$$

得到了 \mathbb{R}^3 中一组漂亮的**正交归一基**。下图的绿色屏幕就是本例的切空间，红色箭头为 ω ，两个蓝色箭头分别为 u, v 。



在古典微分几何中，它一般充当**活动标架**，随着流形上的点 p 运动，每到一处，就建立一个新的标架。这个功能的实用性很强，例如飞机就可以看作一个活动标架，它在地球表面（流形 S^2 上）飞行，而你坐在飞机内，感受到的是这个标架坐标系，而非机场的地面坐标系。

微分几何基本理论

刚才的计算部分，对于工程师来说，应该很熟悉。接下来的理论部分，可能会有些概念比较陌生。但是没关系，请记住诺伊曼的话，就是一些名字而已，熟悉后就好了。让我们从这个结构开始：

$$df = \partial_1 f dx^1 + \partial_2 f dx^2 + \partial_3 f dx^3 = \partial_\mu f dx^\mu$$

在微分几何中， df 的正式名称叫做**微分1-形式**。怎么来的呢？是因为 d 这个符号，在微分几何中，被称作**外微分算子**。 d 作用于一个光滑函数 f ，生成一个 1-形式，即

$$df = \partial_\mu f dx^\mu$$

在这个过程中， f 被称为 **0-形式**，而 df 是一个 **1-形式**，因为它将切空间中的向量映射到实数。由于 df 是通过外微分算子 d 作用于 0-形式 f 得到的，它通常被称为 **微分1-形式**。外微分 d 作用在 1-形式 $\omega = \omega_\mu dx^\mu$ 上，生成 2-形式，写作：

$$d\omega = \partial_\nu \omega_\mu dx^\nu \wedge dx^\mu = \frac{1}{2} (\partial_\nu \omega_\mu - \partial_\mu \omega_\nu) dx^\nu \wedge dx^\mu$$

算子 d 作用在 k -形式上，就会提升一级。如果连续对光滑函数 f 提升两次，就会变成 0，即： $d^2 = 0$ 。

这些微分形式和外微分的概念在理论物理中非常重要，特别是在电磁学中，**梯度**、**旋度**和**散度**都可以用微分形式的语言来表达。比如，电场的 **梯度** 可以用 1-形式表示，磁场的 **旋度** 可以用 2-形式表示，而电荷密度的 **散度** 可以通过 3-形式来描述。

暂不考虑前面的修饰词**微分**，实际上**1-形式**这个概念，初中就接触过了。比如直线方程：

$$ax + by = c$$

我们可以将它理解为一个 1-形式作用在向量上的结果。不信的话，我们按照微分几何的标准，将它用 1-形式的语言重新表达：设 $\omega_1 = a$ ， $\omega_2 = b$ ，则 1-形式为：

$$\omega = a dx^1 + b dx^2 = \omega_1 dx^1 + \omega_2 dx^2 = \omega_\mu dx^\mu, \quad \mu = 1, 2$$

注意，这里用了爱因斯坦求和约定。再设 $v^1 = x$ ， $v^2 = y$ ，则向量为：

$$\mathbf{v} = x \frac{\partial}{\partial x^1} + y \frac{\partial}{\partial x^2} = v^1 \frac{\partial}{\partial x^1} + v^2 \frac{\partial}{\partial x^2} = v^1 \partial_1 + v^2 \partial_2 = v^\nu \partial_\nu, \quad \nu = 1, 2$$

于是 1-形式 ω 作用在向量 \mathbf{v} 就写成：

$$\omega(\mathbf{v}) = ax + by$$

一通操作猛如虎。原本你会的不能再会的初中数学，变成了看不懂的玩意。

$$ax + by = c \quad \text{vs} \quad \omega(\mathbf{v}) = c$$

它的意思是，1-形式 ω 作用在切向量 \boldsymbol{v} 上，返回实数 c 。严格的表达为：

$$\omega : \mathbb{R}^2 \rightarrow \mathbb{R}, \quad \boldsymbol{v} \mapsto \omega(\boldsymbol{v})$$

所以，直线方程 $ax + by = c$ 可以理解为，被1-形式 ω 作用在 \mathbb{R}^2 上的某些点的结果为常数 c ，这些点构成了一条直线。它其实是用1-形式（协变向量）刻画出的等值超平面。

$$l = \{\boldsymbol{v} \in \mathbb{R}^2 \mid \omega(\boldsymbol{v}) = c\}$$

至此你应该对微分1-形式没有那么畏惧了，我们继续学下去，很快你就会发现，微分几何，会从遥不可及的高冷女神，变成人见人爱的邻家小妹。

接下来让我们正本逐源，从线性映射的概念开始：

设 V, W 都是域 \mathbb{F} （一般为 \mathbb{R}, \mathbb{C} ）上的向量空间（纯数学谓之线性空间），如果映射：

$$f : V \rightarrow W$$

对任意 $x_1, x_2 \in V$ ，任意 $a \in \mathbb{F}$ ，若

$$f(x_1 + x_2) = f(x_1) + f(x_2), \quad f(ax_1) = af(x_1)$$

则称 $f : V \rightarrow W$ 是一个**线性映射**。

为聚焦核心概念，我们将 V, W 简化为 $\mathbb{R}^m, \mathbb{R}^n$ 。即：

$$f : \mathbb{R}^m \rightarrow \mathbb{R}^n$$

1. $m = 1, n = 1$: 即 $y = kx$ ，这是最简单的线性映射。

$$f : \mathbb{R} \rightarrow \mathbb{R}, \quad x \mapsto kx, \quad k \neq 0$$

不妨验证一下：

- 加法保持： $f(x_1 + x_2) = k(x_1 + x_2) = kx_1 + kx_2 = f(x_1) + f(x_2)$
- 数乘保持： $f(ax) = k(ax) = a(kx) = af(x)$

你会发现，验证过程，用到了结合律，交换律，分配律。这些都是小学数学必修知识，看来你读

的是布尔巴基小学，能够明白线性空间是一种特殊的模，而线性映射就是模同态映射。

2. $m = 1, n = 2$: 即 $f(x) = (k_1x, k_2x)$, 值域稍做推广。

$$f: \mathbb{R} \rightarrow \mathbb{R}^2, \quad x \mapsto (k_1x, k_2x), \quad k_1, k_2 \neq 0$$

我们马上就会面临，用什么符号，或者什么方式，才能最简洁有效的表达这个线性映射 f 。

我们知道，从 \mathbb{R} 到 \mathbb{R}^2 的线性映射 f 可以看作是由两个从 \mathbb{R} 到 \mathbb{R} 的线性映射组合而成。为了更简洁地表示这种映射关系，数学家们引入了矩阵的概念。对于上述线性映射 f ，其对应的矩阵为 $A = \begin{bmatrix} k_1 \\ k_2 \end{bmatrix}$ 。根据矩阵乘法规则，对于任意的 $x \in \mathbb{R}$,

$$f(x) = Ax = \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} [x] = \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} x = \begin{bmatrix} k_1x \\ k_2x \end{bmatrix}$$

本质上，矩阵形式是通过两个相对独立（其实有协同）的部分来实现从 \mathbb{R} 到 \mathbb{R}^2 的线性映射。这两个部分可以看作是分别作用的线性映射：

$$\begin{aligned} f_1: \mathbb{R} &\rightarrow \mathbb{R}, & x &\mapsto k_1x \\ f_2: \mathbb{R} &\rightarrow \mathbb{R}, & x &\mapsto k_2x \end{aligned}$$

并且按照矩阵乘法的规则进行运算，最终就可以达到想要的结果。

3. $m = 2, n = 1$: 即 $f(x_1, x_2) = k_1x_1 + k_2x_2 = y$, 定义域稍作推广。

$$f: \mathbb{R}^2 \rightarrow \mathbb{R}, \quad (x_1, x_2) \mapsto k_1x_1 + k_2x_2 = y, \quad k_1, k_2 \neq 0$$

同样，我们可以将其表示为矩阵形式：

$$f(x_1, x_2) = \begin{bmatrix} k_1 & k_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = k_1x_1 + k_2x_2 = y$$

请注意与 $m = 2, n = 1$ 与 $m = 1, n = 2$ 的区别。后者有个合并同类项的过程，即：

$$k_1x_1 + k_2x_2 = y$$

这就是问题的关键。什么时候能合并？什么时候不能？为什么？如同政治经济学中的国家合并，企业重组。有独立性就不能合并同类项，合并了就说明丧失了独立性。或者说，不同物种之间有

生殖隔离，才能保障生物多样性。

我们举个不太恰当的例子：你想做一盘沙拉，用了3个香肠，2个洋葱。你想表达食材的构成，要写成：

$$3\text{个香肠} + 2\text{个洋葱}$$

但是你不能写成：

$$(3 + 2)(\text{香肠} + \text{洋葱})$$

这会造成歧义。并且，根据数学运算法则，还会导致：

$$(3 + 2)(\text{香肠} + \text{洋葱}) = 5(\text{香肠} + \text{洋葱}) = 5\text{个香肠} + 5\text{个洋葱}$$

这样搞是魔术，不是数学。原本3个香肠，2个洋葱，变成了5个香肠，2个洋葱。

我们称“香肠”和“洋葱”为**基**，前面的系数“2”，“3”为**分量**。基表明两者是独立的食材，并非同类，当然也就不能合并同类项。因此写出下式不会产生歧义：

$$3\text{ 香肠} + 2\text{ 洋葱}$$

但是，现在我想做沙拉了。于是把洋葱和香肠切碎了，加上了一些其他的食材，混在一起做成沙拉端上桌。这个时候，又认为洋葱和香肠可以合并了，或者消失了，转化成了沙拉。因此，你可以理解为：

$$k_1x_1 + k_2x_2 = y \quad \Longleftrightarrow \quad 3\text{香肠} + 2\text{洋葱} = \text{沙拉}$$

注意到 $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ 的结构，与之前的1-形式作用于切向量结果为常数 c 的结构，两者几乎一模一样。

$$k_1x_1 + k_2x_2 = y \quad \text{vs} \quad ax + by = c$$

在自然界有个现象，长得像就一定有长得像的理由。例如几乎所有的哺乳动物都有四条腿，这是由内在的基因决定的。据此，我们可以笃定，1-形式是一种特殊的线性映射。只不过，1-形式的表达式，同时带着基和分量的信息。而矩阵表达，只有分量，没有基的信息。

这是两种风格，就好比业余棋手下的棋，与职业对局的区别。业余棋手可能靠直觉选点、靠计算

吃子；职业棋手却是基于对全局形势判断（外势和实空）的前提下追求每个子的效率。

为何只有分量的矩阵运算能够正常运转呢？我们必须要在此刻导入对偶空间的概念，才能完全理解这件事情。

设 V 是一个向量空间，其元素为向量。**对偶空间** V^* 是由 V 上的一类特定的**线性映射**构成的集合，这些线性映射的作用是将向量空间 V 中的元素映射到标量场（通常是 \mathbb{R} 或 \mathbb{C} ）。此时我们将这种类型的线性映射称为**线性形式**，

$$V^* = \{f : V \rightarrow \mathbb{R} \mid f \text{ 是线性映射}\}.$$

对偶空间有两个核心属性，分别对偶基与线性性质：

1. **基与对偶基**：如果 V 有一组基 $\{e_1, e_2, \dots, e_n\}$ ，那么对偶空间 V^* 中有一组对偶基 $\{e^1, e^2, \dots, e^n\}$ ，其中：

$$e^i(e_j) = \delta_{ij}$$

其中 δ_{ij} 是克罗内克 delta 函数， $\delta_{ij} = 1$ 当 $i = j$ ，否则 $\delta_{ij} = 0$ 。

2. **线性性质**：对偶空间 V^* 中的元素是线性形式 f ，而 f 本身也满足线性性质，故也可被视为向量，一般称**对偶向量**以示区别。线性性质体现为，对任意 $f_1, f_2 \in V^*$ ， $v \in V$ ， $a \in \mathbb{F}$ ：

- 加法保持 $(f_1 + f_2)(v) = f_1(v) + f_2(v)$
- 数乘保持 $(af_1)(v) = a(f_1(v))$

对偶空间 V^* 本质上是由作用于 V 的所有线性形式构成的集合。每个向量空间 V 都有一个对应的对偶空间 V^* ，且如果 V 是有限维的， V 和 V^* 之间是同构的，这意味着对偶空间 V^* 自身也是一个向量空间。

继续下一步之前，我们需要解决掉一些术语滥用的小麻烦。下表是各种术语之间，细微的差别（不考虑复数域 \mathbb{C} ）：

术语	数学表达式	所属领域	简要说明
线性映射	$f : V^m \rightarrow W^n$	线性代数	通用术语，可用于任意两个向量空间之间

术语	数学表达式	所属领域	简要说明
线性形式	$f : V^m \rightarrow \mathbb{R}$	线性代数	通常用于有限维
线性函数	$f : \mathbb{R}^n \rightarrow \mathbb{R}$	线性代数	通常指欧氏空间到实数域
线性泛函	$f : V \rightarrow \mathbb{R}$	泛函分析	通常用于无穷维，强调是函数空间的“函数”
1-形式	$\alpha : T_p M \rightarrow \mathbb{R}$	微分几何	$\alpha \in T_p^* M$ ，即切空间的对偶；强调与坐标无关
微分1-形式	$\omega : T_p M \rightarrow \mathbb{R}$	微分几何	$\omega = df$ ， 强调可由外微分得到，又叫恰当形式

一开始理解对偶空间会有些困难，最好的方式就是，首先相信前辈数学家的结论，其次就是找些具体的例子来验证。当我们看到等式

$$ax + by = c$$

用矩阵表示为：

$$A\boldsymbol{v} = c$$

即：

$$A = \begin{bmatrix} a & b \end{bmatrix}, \quad \boldsymbol{v} = \begin{bmatrix} x \\ y \end{bmatrix}$$

则：

$$A\boldsymbol{v} = \begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = ax + by = c$$

其实我们早就知道，更多的时候，矩阵代表了一个线性映射，而非一组向量。现在有了对偶空间的概念，那么就更容易搞清楚是怎么回事了。

1. 矩阵表示一个线性映射

A 作为一个线性映射，也就是对偶空间 V^* 的元素。完整的表达A，应该带上对偶基：

$$A = ae^1 + be^2$$

而 v 作为线性空间 V 的元素，本身有一组基，因此完整的表达 v ，就应该带上基：

$$v = xe_1 + ye_2$$

于是有：

$$A(v) = Av = (ae^1 + be^2)(xe_1 + ye_2)$$

希望到这里你还没糊涂，刚刚才讲过，线性映射本身具有线性性质，它是加法保持的（见前面对偶空间的定义）。因此有：

$$\begin{aligned} Av &= (ae^1 + be^2)(xe_1 + ye_2) \\ &= ae^1(xe_1 + ye_2) + be^2(xe_1 + ye_2) \end{aligned}$$

而 A 本身就是个线性映射，它作用于向量的时候，原本就是加法保持的（见更前面线性映射的定义）。因此继续展开就有：

$$\begin{aligned} Av &= ae^1(xe_1 + ye_2) + be^2(xe_1 + ye_2) \\ &= ae^1(xe_1) + ae^1(ye_2) + be^2(xe_1) + be^2(ye_2) \\ &= axe^1(e_1) + aye^1(e_2) + bxe^2(e_1) + bye^2(e_2) \end{aligned}$$

此时再看刚才的对偶基定义：

$$e^i(e_j) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

因为 $e^1(e_2) = 0$, $e^2(e_1) = 0$, 于是：

$$Av = axe^1(e_1) + aye^1(e_2) + bxe^2(e_1) + bye^2(e_2) = ax + by$$

也即是说，如果矩阵 A 为对偶空间 V^* 中的元素，也就是一个线性形式，而非向量时。其作用在 V 中的向量 v 上，基和对偶基自动结合后消消乐了，最后的结果是个实数。

这没有什么神奇的地方，因为线性形式就是这么定义的，按定义它就是对偶空间的元素，就自带了对偶基。正如徐志摩所写：轻轻的我走了，正如我轻轻的来。我挥一挥手，不带走天边的一片

云彩。

概念	所属空间	角色	示例
列向量 \boldsymbol{v}	V	原空间中的向量	$\boldsymbol{v} = xe_1 + ye_2$
行向量 A	V^*	对偶空间中的线性形式	$A = ae^1 + be^2$
矩阵乘法 $A\boldsymbol{v}$	$V^* \times V \rightarrow \mathbb{R}$	自然配对操作	$A\boldsymbol{v} = ax + by$

如果我们理解了这一点，很容易推广到一般性的m维到n维的线性映射，也用矩阵来表达：

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}, \quad \boldsymbol{v} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

则有：

$$A\boldsymbol{v} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} a_{11}v_1 + a_{12}v_2 + a_{13}v_3 \\ a_{21}v_1 + a_{22}v_2 + a_{23}v_3 \end{bmatrix}$$

在上面的例子中， $A \in V^*$ ， $\boldsymbol{v} \in V$ ， $A\boldsymbol{v} \in W$ ，映射为：

$$A : V \rightarrow W, \quad \boldsymbol{v} \mapsto A(\boldsymbol{v})$$

其中 $\dim V = 3$ ， $\dim W = 2$ 。

看的很清楚， A 的每一行是一个线性形式，分别将 \boldsymbol{v} 映射为标量。由于矩阵的结构天然有次序，因此可以认为第一行映射得到的标量是第一个分量，第二行映射出第二个分量。这两个分量构成 W 空间中的2维向量。

还可以进一步推广， A 依次映射一组多个向量，例如：

$$\begin{aligned}
AB &= \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} u_1 & v_1 \\ u_2 & v_2 \\ u_3 & v_3 \end{bmatrix} \\
&= \begin{bmatrix} a_{11}u_1 + a_{12}u_2 + a_{13}u_3 & a_{11}v_1 + a_{12}v_2 + a_{13}v_3 \\ a_{21}u_1 + a_{22}u_2 + a_{23}u_3 & a_{21}v_1 + a_{22}v_2 + a_{23}v_3 \end{bmatrix} \\
&= \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \end{bmatrix}
\end{aligned}$$

A 将 B 中存储的两个 3 维向量 \mathbf{u}, \mathbf{v} 分别映射为 2 维向量 \mathbf{x}, \mathbf{y} 。

在这个例子中，虽然 A, B 长的样子都是矩阵，但是其代表的数学对象截然不同。 A 的行属于 V^* ，即 $A = [\alpha, \beta]^\top$ ， $\alpha, \beta \in (V^*)$ ，而 B 的列属于 V ，即 $B = [\mathbf{u} \ \mathbf{v}]$ ， $\mathbf{u}, \mathbf{v} \in V$ 。

矩阵是极其强大并方便直观的数学对象，最适合用于数值计算。因为用索引遍历数据这种事情，人类会觉得比较枯燥和繁琐，而电脑恰恰最擅长干这个。不过要注意一个陷阱，除非你知道矩阵中数据的含义，它代表了什么，才能放心使用。而在微分几何的理论推导中，张量才是真正具有坐标无关性的结构性工具。下面我们马上会谈为什么。

2. 矩阵表示向量内积：

中学阶段，同学们对于 \mathbb{R}^n 的点积就很熟悉了，其定义如下：

$$\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\| \|\mathbf{v}\| \cos \theta$$

其中 θ 是两个向量之间的夹角。如果在坐标系中，实际计算内积。令：

$$\mathbf{u} = u_1 \vec{i} + u_2 \vec{j}, \quad \mathbf{v} = v_1 \vec{i} + v_2 \vec{j}$$

则有：

$$\begin{aligned}
\mathbf{u} \cdot \mathbf{v} &= (u_1 \vec{i} + u_2 \vec{j}) \cdot (v_1 \vec{i} + v_2 \vec{j}) \\
&= u_1 v_1 (\vec{i} \cdot \vec{i}) + u_1 v_2 (\vec{i} \cdot \vec{j}) + u_2 v_1 (\vec{j} \cdot \vec{i}) + u_2 v_2 (\vec{j} \cdot \vec{j}) \\
&= u_1 v_1 + u_2 v_2
\end{aligned}$$

在 \mathbb{R}^n 中，这样讲，虽然不准确，却也没毛病。或许在微分几何出来之前，古典向量分析就是这么干的也或未可知。你可以看到，整个过程，看上去和线性形式作用在向量上的效果一模一样。

这边也可以定义：

$$\vec{i} \cdot \vec{j} = \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

这样就山寨了一套对偶空间的符号系统，运行起来效果杠杠的。不过，山寨货毕竟比不上原装正版。我们还是来学习微分几何语境下，内积（点积）到底是怎么回事。

首先，点积这个说法，被升级成了**内积**，符号表达也变了，微分几何一般写成 $g(u, v)$ ，泛函分析喜欢写成 $\langle u, v \rangle_g$ ，其中 g 被称作**度量张量**。注意，向量 u, v 不再写成粗体了。

虽然不是必须的，但是倘若掌握了张量概念，很自然就能理解内积。而且，现代微分几何，实际上使用张量语言描述。因此，该来的总是躲不过，不如迎难而上，一鼓作气拿下最后一座天王山。

张量简介

从纯数学角度看，张量是一种蛮复杂的数学结构，有点类似多项式，而且每一项的内部结构也挺丰富。我们只研究单项张量，这需要从线性形式开始。先写一个最规范的线性形式，使用希腊字母作为索引指标，以及爱因斯坦求和约定：

$$\omega = \omega_\mu dx^\mu, \quad \mu = 1, 2, \dots, n$$

其中 $\omega \in V^*$ 。实际上这就是个张量，被称为 $(1, 0)$ 型张量，也被称为**1阶协变张量**。这是张量的基本单元之一，你可以把它比作男人。接着再写出一个切向量，依然采用微分几何规范：

$$v = v^\nu \partial_\nu$$

这也是个张量，被称作 $(0, 1)$ 型张量，也被称作**1阶逆变张量**。这是张量的另一个基本单元，你可以把它比作女人。

一共就只有这两种基本的张量，初学者看到张量心生畏惧，多半有两点原因：一是出现了希腊字母。众所周知，在数学，尤其是理论物理领域，如果出现了希腊字母，那就意味着知识层级往上抬了一级。因为初等知识中的公式，已经把拉丁字母用光了。

其次，这个协变基 dx^μ ，逆变基 $\partial_\nu (\frac{\partial}{\partial x^\nu})$ 实在让人一头雾水。实际上这只是个符号而已，只是

微分几何的规范要用这个符号。代数学家更倾向于用 e^i 表示对偶基（协变基）， e_i 表示向量基（逆变基）。实际上这个符号，与你熟悉的 \vec{i}, \vec{j} ，好像也没啥区别，就是个符号而已。

当然了，用这个符号，肯定也是前辈几何学家，权衡再三的结果，的确它非常强烈的暗示了微分结构。 dx^μ 是余切空间的基向量，作用于切空间的向量上能得到有意义的标量结果； $\frac{\partial}{\partial x^\nu}$ 是切空间的基向量，常用来表示向量场在坐标方向上的变化率。

我们有了张量的两种基本单元之后，接下来就是两种基本的运算，一种是张量积，另一种是张量缩并。你大致可以理解为盖房子和拆房子。

1. 张量积

有了基本的(0,1)型和(1,0)型张量，就可以通过张量积来构造高阶张量：

$$A_a \otimes B^b = C_a^b$$

这是彭罗斯发明的指标法，抽象指标用拉丁字母如 a, b, c ，仅用于描述张量的结构和运算逻辑，不涉及具体数值；其中字母严格按字典排列表示槽位次序，上下标严格表示该槽位是向量（切向量）还是对偶向量（线性形式）。当引入坐标系进行实际计算时，需将抽象指标替换为希腊字母作为分量指标，此时每个指标对应着在该坐标系下的具体分量值：

$$u_\mu dx^\mu \otimes v^\nu \partial_\nu = w_\mu^\nu dx^\mu \otimes \partial_\nu$$

如果想继续升阶，就继续通过张量积构造。整个过程就像搭积木，很简单：

$$C_a^b \otimes D^c \otimes E_d^e = T_{ad}^{bce}$$

我们平时看到的矩阵形式的张量表达，一般认为是流形嵌入后的环境坐标，即 \mathbb{R}^n 下的正交归一基，因而可以忽略基，只表达分量：

$$u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}, \quad v = \begin{bmatrix} v^1 \\ v^2 \\ v^3 \end{bmatrix}$$

此时因为涉及具体运算，则抽象指标要改成分量指标，也就是用希腊字母来表达：

$$u_\mu \otimes v^\nu = \mathbf{u} \mathbf{v}^\top = (w_\mu^\nu) = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \begin{bmatrix} v^1 & v^2 & v^3 \end{bmatrix} = \begin{bmatrix} u_1 v^1 & u_1 v^2 & u_1 v^3 \\ u_2 v^1 & u_2 v^2 & u_2 v^3 \end{bmatrix}$$

实际上，用矩阵来表达张量，仅对初学时建立感性认识有帮助。熟悉了之后，还是应该用彭罗斯的抽象指标/分量指标来表达。分析结构，理论推导时，用拉丁字母表示张量类型和结构。要进行分量运算时用希腊字母，而矩阵总是只表示分量。且出现矩阵，就表示要进行实际的运算了。

因此要注意， u_a 表示一阶协变张量， u^b 表示一阶逆变向量，指标 a, b 时表示张量的类型，而非分量索引。也就是，永远不能对拉丁字母赋值，诸如 $a = 1, 2, 3, \dots$ ，这是不合法的。如果仅作理论推导，则应写成：

$$w_a^b = u^a \otimes v^b$$

反之，如果写成 u_μ, v^ν 形式，则说明要对分量运算了，紧跟着就应该出现 $\mu = 1, 2, 3, \dots, n$ 的遍历过程，或者出现矩阵形式。

有了以上感性认识，我们来看看张量的定义：

设 V 是一个 n 维向量空间（定义在域 \mathbb{F} 上，如 \mathbb{R} 或 \mathbb{C} ）， V^* 是其对偶空间（所有线性泛函的集合）。

- 一个 **(p,q) 型张量** T 是一个多重线性映射：

$$T : \underbrace{V^* \times \dots \times V^*}_p \times \underbrace{V \times \dots \times V}_q \rightarrow \mathbb{F}$$

即 T 对每个变量都是线性的。所有 (p, q) 型张量构成的集合记作 $T_q^p(V)$ 。

我们刚才构造出来的张量，其所属的空间就是这样：

$$T_{ad}^{bce} \in V^* \times V \times V \times V^* \times V$$

通过张量积构造高阶张量的过程，你可以理解为，你手上有饺子皮，饺子馅若干。你根据需要，将它们排成了一列，形成一个结构，然后什么也不做，就干等着。

2. 张量缩并

与张量积反过来的操作就是缩并，这是张量真正强大和灵活的地方。由于我们只在平凡的 R^n 中处理张量，因此随便找一个刚才构造的张量，显式完整表达就是：

$$T_{ad}^{bce} = T_{ad}^{bce} dx^a \otimes \partial_b \otimes \partial_c \otimes dx^d \otimes \partial_e$$

可能一开始你觉得用 $abcde$ 作为索引很古怪，但是没办法，高阶张量的经常超过3阶， ijk 作为索引不够用，而且 ijk 有可能暗示向量空间是 R^n 。张量灵活的地方有两个：其一，它既可以作用于向量（切向量，逆变向量），也可以作用于对偶向量（微分形式，线性形式，协变向量，余切向量）。其二， (p, q) 型张量，可以一次性的作用最少1个最多 p 个向量，也可以作用最少1个最多 q 个对偶向量，只要位置对应上即可：

$$\begin{aligned} T_{ad}^{bce} u^a v_c &= T_{ad}^{bce} (u^a, -, v_c, -, -) = (T_{ad}^{bce} dx^a \otimes \partial_b \otimes \partial_c \otimes dx^d \otimes \partial_e) (u^a \partial_a, -, v_c dx^c, -, -) \\ &= (T_{ad}^{bce} u^a v_c) (dx^a (\partial_a) \otimes \partial_b \otimes dx^c (\partial_c) \otimes dx^d \otimes \partial_e) \\ &= S_d^{be} \partial_b \otimes dx^d \otimes \partial_e \end{aligned}$$

这就是张量的缩并，如果只看分量的变化，那么按照爱因斯坦求和约定无脑运算即可，重复的上下指标，就通过求和合并了。注意上式中：

$$dx^a (\partial_a) = \delta_a^a = 1, \quad dx^c (\partial_c) = \delta_c^c = 1$$

这里的 δ_a^b 抽象的克罗内克符号，不是用于分量计算的克罗内克符号。 a, b 匹配时返回的0/1 应理解为匹配上就返回1，否则返回0。

$$\delta_a^b = \begin{cases} 1 & a = b \\ 0 & a \neq b \end{cases}$$

如果你真的理解了线性代数，张量不过时推广到了多线性代数。不要被上面一大串吓住了，我们从几个地方来尝试理解张量，保证让你觉得不过如此。

首先，张量的构造过程，没有什么好说的，一共就两种类型。来了男人就把男人加进这个社区，来了女人就把女人加进这个社区。然后就没有然后了。

它的灵活性体现在缩并。缩并的时候，完全不一样，必须男女严格配对。协变槽位，作用在一个逆变向量上。逆变槽位，作用在一个协变向量上。同时它又是男女平等的，并无主从之分。且不限于谁主动，谁先发起，女上位也是可以的。

总结下来就是：构造是自由开放的，缩并要求男女（女男）搭配。

✓ 1阶张量术语对照表（协变/逆变类型）

名称	类型	空间	常用领域	备注说明
协变向量	(1,0)	V^*	广义相对论	语义强调变换性，与“坐标变换下协变”相关
对偶向量	(1,0)	V^*	线性代数	强调与向量空间 V 的配对关系
余切向量	(1,0)	V^*	微分几何	强调作用在切空间上的对象
线性形式	(1,0)	V^*	线性代数	从函数视角： $\phi : V \rightarrow \mathbb{R}$
1-形式	(1,0)	V^*	微分几何	是 1 阶微分形式，可全局定义在流形上
微分1-形式	(1,0)	V^*	微分几何	外微分df生成的1-形式
逆变向量	(0,1)	V	广义相对论	表示随坐标变换“反向变化”的对象
切向量	(0,1)	V	微分几何	在流形中表示方向导数或速度向量

写到这里，我感觉术语滥用又比较严重了，因此列一个表出来。看来鲁迅说的没错，所谓学问，不过是知道茴字有几种写法。

度量张量与内积

对张量有了初步了解之后，我们来看看度量张量g。

- **度量张量 g** ：一个对称、正定的 $(0, 2)$ 型张量，一般也叫二阶协变张量，用于定义内积：

$$g(X, Y) = \langle X, Y \rangle_g \quad X, Y \in T_p M$$

在局部坐标下写成这样：

$$g = g_{ij} dx^i \otimes dx^j$$

- **逆度量张量 g^{-1}** ：满足 $g^{ik} g_{kj} = \delta^i_j$ 的 $(2, 0)$ 型张量，也叫二阶逆变张量，长成这样：

$$g^{-1} = g^{ij} \frac{\partial}{\partial x^i} \otimes \frac{\partial}{\partial x^j} = g^{ij} \partial_i \otimes \partial_j$$

除此之外，还有较少见的混合度量张量，即(1,1)型张量。以下是三类度量张量的对比表：

类型	名称	分量	基	分量作用	几何操作
(2,0)	二阶协变张量	g_{ij}	$dx^i \otimes dx^j$	$g(\mathbf{u}, \mathbf{v}) = g_{ij}u^i v^j$	降指标
(0,2)	二阶逆变张量	g^{ij}	$\partial_i \otimes \partial_j$	$g^{-1}(\alpha, \beta) = g^{ij}\alpha_i \beta_j$	升指标
(1,1)	二阶混合张量	$g^i_j = \delta^i_j$	$\delta^i_j \partial_i \otimes dx^j$	$g^i_j v^j = v^i$	指标平衡

• **内积** $g(u, v)$ ： 一个 $(2, 0)$ 型张量，用于度量两个向量之间的距离。

$$g(u, v) = \langle u, v \rangle_g = g_{ab}u^a v^b \quad u, v \in T_p M$$

于是，你看到的两个向量内积，实际是是过度规张量，作用在第一个向量上，返回一个线性形式,这个过程称之为**降指标**。然后再用返回的线性形式，作用在第二个向量，最后返回一个标量：

$$g(u, v) = g_{ab}u^a v^b = (g_{ab}u^a)v^b = \tilde{g}_b v^b = \tilde{g}_b(v^b)$$

可能还是需要举个实际的例子: 流形 M 上的度量张量为 g_{ab} ，给定流形上一点 p ，在坐标系 x^μ 下，点 p 处的度量张量的分量为 $g_{\mu\nu}$ ，排列称矩阵格式 $(g_{\mu\nu})$ ：

$$g = (g_{\mu\nu}) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

点 p 的切空间 $T_p M$ 上的量个向量 u^a, v^b 在坐标系 x^μ 下的分量为 u^μ, v^μ ：

$$u = (u^\mu) = \begin{bmatrix} 3 \\ 2 \\ 7 \end{bmatrix}, \quad v = v^\mu = \begin{bmatrix} 4 \\ -5 \\ 1 \end{bmatrix}$$

注意，我尽量用严谨的微分几何术语及符号，来表达一切。最后出现的 g, u, v 相当于写程序的变量名，已经脱离了张量的命名规范。在此之前，所有的命名和符号，都符合微分几何规范。

则点 p 处的内积 $g(u, v)$ 计算过程如下：

$$\begin{aligned}
g(u, v) &= g_{\mu\nu} u^\mu v^\nu = \sum_{\nu=1}^3 \sum_{\mu=1}^3 g_{\mu\nu} u^\mu v^\nu = \sum_{\nu=1}^3 \left(\sum_{\mu=1}^3 g_{\mu\nu} u^\mu \right) v^\nu \\
&= \sum_{\nu=1}^3 (g_{1\nu} u^1 + g_{2\nu} u^2 + g_{3\nu} u^3) v^\nu \\
&= g_{11} u^1 v^1 + g_{22} u^2 v^2 + g_{33} u^3 v^3 \\
&= 1 \times 3 \times 4 + 2 \times 2 \times (-5) + 3 \times 7 \times 1 \\
&= 12 - 20 + 21 = 13
\end{aligned}$$

注意，本来展开后一共有9项，但是非对角线项为0，因此只需要计算对角线项即可。上述计算过程，写成矩阵形式就是：

$$\begin{aligned}
g(u, v) &= guv = (gu)^T v = u^T g^T v = u^T gv \\
&= \begin{bmatrix} 3 & 2 & 7 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} 4 \\ -5 \\ 1 \end{bmatrix} = 13
\end{aligned}$$

注意对称张量有个特点： $g^T = g$ 。

因此，当你理解了张量之后，再看到诸如 $u^T gv$ 这样的表达式，就会对转置很敏感。其实它暗示了把向量 u 作用在 g 上，返回一个线性形式，然后再作用在向量 v 上，返回一个标量。

因此我们得出一个结论，线性形式作用在向量上，返回一个标量，这个才是最基本的操作。而内积则很复杂，需要翻来覆去的折腾。在 R^n 中，我们感受不到这种折腾，是因为 R^n 是一个平凡的线性空间，它度量张量就是单位矩阵。例如 R^3 中的度量张量就是：

$$g = \mathbf{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

因此， R^3 中的内积就是向量的点积，也就是我们熟悉的：

$$g(u, v) = u^T \mathbf{I} v = u^T v = u \cdot v$$

此时，度量张量 g 对 u 的分量降指标，过程悄无声息，因为分量的值没有任何改变，只是改变了 u 的性质。使得整件事情，在整个代数理论框架下能够自圆其说。

前面对张量的缩并过程的讲解，虽然没有什么谬误，但是疏漏之处颇多，很多概念也不够清楚。实际上，张量的缩并，并非只能理解为：按槽位吃进一个向量进行缩并，它是一种很灵活的操作。我们用内积的例子来体会一下：

$$g(u, v) = g_{ab} u^a v^b = (g_{ab} u^a) v^b = \tilde{g}_b v^b$$

之前这样理解，并无问题。但还可以用另一种方式来理解：

$$g(u, v) = g_{ab} (u^a \otimes v^b) = g_{ab} w^{ab}$$

其中 $w^{ab} = u^a \otimes v^b$ 是 a 和 b 的张量积。也就是说，它无所谓 u, v 是否先做了张量积。 g 去缩并的时候，只要 w^{ab} 中的 a 和 b 匹配上即可。我们来验证一下是否果真如此：

$$\begin{aligned} w &= (w^{\mu\nu}) = u^\mu \otimes v^\nu = uv^T \\ &= \begin{bmatrix} 3 \\ 2 \\ 7 \end{bmatrix} \begin{bmatrix} 4 & -5 & 1 \end{bmatrix} = \begin{bmatrix} 12 & -15 & 3 \\ 8 & -10 & 2 \\ 28 & -35 & 7 \end{bmatrix} \end{aligned}$$

那么：

$$\begin{aligned} (g_{\mu\nu} w^{\nu\mu}) &= g_{\mu\nu} w^{\mu\nu} = gw \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} 12 & -15 & 3 \\ 8 & -10 & 2 \\ 28 & -35 & 7 \end{bmatrix} = \begin{bmatrix} 12 & -15 & 3 \\ 16 & -20 & 4 \\ 84 & -105 & 21 \end{bmatrix} \end{aligned}$$

好像不太对，并没有得到我们想要的结果，标量值 13，出来的还是个矩阵。但是别着急，注意看结果矩阵的对角线元素：12, -20, 21，把三者相加：

$$12 + (-20) + 21 = 13$$

这就是我们想要的结果，内积 $g(u, v)$ 的值。因此，如果用矩阵来表示张量，并进行缩并运算，

那么需要根据两种情况来分别处理：

$$g(u, v) = (g_{\mu\nu} u^\mu v^\nu) = (gu)^T v = u^T g^T v$$

另一种是：

$$g(u, v) = g_{\mu\nu} (u^\mu \otimes v^\nu) = g_{\mu\nu} w^{\mu\nu} = \text{Tr}(gw)$$

此时gw依然是个方阵，求它的迹，也就是主对角线元素之和，才是完全缩并后的结果。这个运算，在python的numpy库中，就是用 `tensor_dot` 函数来实现的。

```
# 使用tensor_dot显式指定缩并槽位进行完全缩并
result = np.tensor_dot(g, w, axes=([0,1],[0,1]))
```

如果不想封装，想看先矩阵相乘，再对结果矩阵求迹的过程：

```
# 按照矩阵乘法规则计算 g·w
gw = np.dot(g, w)

# 计算矩阵gw的迹
result = np.trace(gw)
```

两种办法都可以得到完全缩并后的结果。从这个例子可以看出，张量的缩并，虽然实际的计算过程，的确是一男一女匹配上，然后就消消乐了。但是从形式上，可以是任意两个张量之间缩并，只要精确指定索引匹配上即可。即：

$$S_{ab}^{cd} T_c^{be} = W_a^{de}$$

可见张量积的确什么都没敢，他就是数据的搬运工，制定了数据的摆放方式，建好档案，然后就不管了。将来要用到的时候，再按照摆放方式，把数据拿出来。你也可以理解为，张量积就是两拨人合作工作的方式，他们之间没有任何关系。男人和男人可以一起工作，女人和女人也可以一起工作，男人女人也可以一起工作。

但是缩并不同，上面的例子，好比 S 家族一群人有男有女（指标有上有下），与 T 家族另一群

人，开了个相亲大会。规矩就是，家族之内的人，不能互相谈恋爱。只能 ST 两个家族之间的人谈恋爱。还有一条规矩：不允许男人之间搞对象，也不允许女人之间搞对象，必须是男女搭配。一旦互相对上眼了，男女两人就牵手，退出相亲大会，如上例配对成功 b, c 两对四人。剩下的单身狗，通过张量积形成新张量的索引，等待下一次相亲大会。

★ 张量积=工作会议 vs 缩并=相亲大会：

操作	本质	比喻
张量积	无互动，仅组装高阶结构	两拨人建档合作
张量缩并	索引配对，相同指标消失	相亲会，男女牵手
内积	特例：度规缩并两个向量	直接婚配成标量
索引留下	没被配对的，形成新张量的索引	单身狗继续开会

✓ 一般不要求约束流形是“闭的流形”，但常常要求它是闭集 (closed set)

这是一个 subtle（微妙）但非常重要的区分：

◆1. “闭流形”与“闭集”是两个概念：

- 闭流形 (closed manifold)：是一个无边界的紧流形（即紧且无边界的流形），例如 S^1 , S^2 （单位圆、单位球面）。
- 闭集 (closed set)：是指它在所嵌入的拓扑空间中是闭的，例如在 \mathbb{R}^n 中包含其所有极限点。

例子对比：

- $y = x^2$ 是个光滑流形，但它不是紧的（因为 $x \in \mathbb{R}$ 无界），所以不是“闭流形”，但作为 \mathbb{R}^2 中的一个集合，它确实是闭集。
- $S^1 \subset \mathbb{R}^2$ 是一个闭流形，也当然是闭集。

◆2. 在凸优化中，我们关心的是可行域是闭集，不一定是“闭流形”：

在约束优化中，比如带有等式约束：

$$\text{minimize } f(x) \quad \text{subject to } h_i(x) = 0$$

我们考虑的约束集通常是形如

$$M = \{x \in \mathbb{R}^n \mid h_1(x) = \cdots = h_k(x) = 0\}$$

只要这些 h_i 足够光滑，并且雅可比矩阵 $Dh(x)$ 满秩，**根据隐函数定理**，这个集合局部是一个光滑流形。

✓ 然而，我们不会要求这个集合是“闭流形”，但通常我们会希望：

- 这个集合在 \mathbb{R}^n 中是**闭集**（以保证极小值存在）；
- 最好是**紧集**（紧+闭），这样通过**极值定理**可以保证最小值取得。

◆3. 所以实务中我们更关心的是：

条件	是否常要求	原因
光滑流形结构	✓ 是	为了定义梯度、拉格朗日乘子、KKT 条件等
闭集	✓ 是	确保极限点属于可行域，便于存在性分析
紧集	✓ 常常希望	保证极值存在
闭流形（紧+无边界的流形）	✗ 一般不要求	太强了，不适用于很多实际模型

✓ 小结一句话：

在实际的凸优化中，约束集合常常是一个**嵌入在 \mathbb{R}^n 中的光滑子流形**，我们通常希望它是一个**闭集**，但不要求它是“闭流形”。

如果你是从最优化中的拉格朗日乘子法、约束空间内的梯度下降、投影法等角度思考这个问题，这种对“闭集”与“闭流形”的区分是非常有价值的！

你有在看具体哪一类优化问题吗？比如约束在球面上？在矩阵流形上？还是机器学习里的？我可以结合实例给你解释。

2. D 是 \mathbb{R}^2 中的一个闭圆盘：

这种情况下，我们可以直接对 $f(x, y)$ 求偏导，得到 $f_x(x, y), f_y(x, y)$ ，然后分别对 x, y 求导，找到驻点。

注意：

2. 引入拉格朗日乘子 $\alpha_i \geq 0$ ，将约束融入目标：

$$\mathcal{L}(\mathbf{w}, b, \alpha) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^m \alpha_i [y_i(\mathbf{w}^\top \mathbf{x}_i + b) - 1]$$

3. 求导并得KKT条件

对 \mathbf{w} 和 b 求偏导，令导数为零：

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^n \alpha_i y_i \mathbf{x}_i = 0 & \Rightarrow \mathbf{w} = \sum_{i=1}^n \alpha_i y_i \mathbf{x}_i \\ \frac{\partial \mathcal{L}}{\partial b} = - \sum_{i=1}^n \alpha_i y_i = 0 & \Rightarrow \sum_{i=1}^n \alpha_i y_i = 0 \end{cases}$$

4. 转化为对偶问题

将 \mathbf{w} 表达式代入拉格朗日函数，消去原变量，得到对偶问题：

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \mathbf{x}_i^\top \mathbf{x}_j \\ \text{s.t.} \quad & \sum_{i=1}^n \alpha_i y_i = 0, \quad \alpha_i \geq 0 \ (\forall i) \end{aligned}$$

5. 求解对偶问题

使用二次规划算法（如SMO算法）求解 α ，得到非零 α_i 对应的样本点——这些点称为**支持向量**，位于间隔边界上（满足 $y_i(\mathbf{w}^\top \mathbf{x}_i + b) = 1$ ）。

6. 计算超平面参数

• **法向量**：由支持向量加权求和得到

$$\mathbf{w} = \sum_{i \in \text{SV}} \alpha_i y_i \mathbf{x}_i$$

• **截距项**：任选一个支持向量 \mathbf{x}_k 计算

$$b = y_k - \mathbf{w}^\top \mathbf{x}_k$$

7. 最终超平面方程

$$\mathbf{w}^\top \mathbf{x} + b = 0 \quad \text{或} \quad \sum_{i \in \text{SV}} \alpha_i y_i \mathbf{x}_i^\top \mathbf{x} + b = 0$$

关键特性与验证

1. **支持向量决定超平面**：仅支持向量对应的 $\alpha_i > 0$ ，非支持向量对 \mathbf{w} 无贡献
2. **间隔验证**：间隔宽度为 $2/\|\mathbf{w}\|$ ，可通过计算几何间隔确认

3. **分类正确性**: 所有样本满足 $y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1$

示例演示（二维简化）

假设数据点:

- 正类: $\mathbf{x}_1 = (1, 1), y_1 = 1$
- 负类: $\mathbf{x}_2 = (2, 2), y_2 = -1$

求解步骤:

- 对偶问题解得 $\alpha_1 = \alpha_2 = 0.5$
- 法向量 $\mathbf{w} = 0.5 \cdot 1 \cdot (1, 1) + 0.5 \cdot (-1) \cdot (2, 2) = (-0.5, -0.5)$
- 截距 $b = 1 - (-0.5 \cdot 1 - 0.5 \cdot 1) = 2$
- 超平面方程: $-0.5x_1 - 0.5x_2 + 2 = 0$ 或 $x_1 + x_2 = 4$

总结: 硬间隔SVM通过凸优化理论保证解的唯一性和全局最优性, 支持向量揭示了数据的关键分类信息。

问题建模

我们致力于构建一个智能分类系统, 能够根据植物形态特征自动判别鸢尾花种类。本案例研究聚焦于二分类任务, 将鸢尾花样本准确划分为山鸢尾 (标签0) 与变色鸢尾 (标签1)。

数据架构

样本空间:

- 总样本量: 100个平衡样本 (山鸢尾与变色鸢尾各50个)
- 特征空间: \mathbb{R}^4 , 包含形态测量数据

$$\mathbf{x} = [x_{\text{sepal_len}}, x_{\text{sepal_wid}}, x_{\text{petal_len}}, x_{\text{petal_wid}}]^\top$$

- 标签空间: $\mathcal{Y} = \{0, 1\}$, 经植物学家权威标注

2. 线性 SVM 的优化问题

SVM 的目标是找到一个超平面：

$$w^T \mathbf{x} + b = 0$$

使得数据点尽可能被正确分类，并且间隔（margin）最大。即：

$$y_i(w^T \mathbf{x}_i + b) \geq 1, \quad \forall i$$

其中：

- $w \in \mathbb{R}^d$ 是法向量，决定超平面的方向。
- $b \in \mathbb{R}$ 是偏置项，决定超平面的位置。

最优化目标：

$$\min_{w,b} \frac{1}{2} \|w\|^2$$

约束条件：

$$y_i(w^T \mathbf{x}_i + b) \geq 1, \quad \forall i$$

这就是 **硬间隔 SVM (hard-margin SVM)**，它假设数据完全线性可分。

监督学习范式

我们采用**有监督学习框架**：

1. **训练阶段**：模型通过解析特征-标签配对数据 $\{(\mathbf{x}_i, y_i)\}_{i=1}^{100}$ 学习判别规律
2. **推理阶段**：对新增样本 \mathbf{x}_{new} 输出预测标签 $\hat{y} \in \{0, 1\}$

关键优势：利用先验知识引导模型学习，显著降低假设空间复杂度。

支持向量机解决方案

在给定特征空间中，我们构建最大间隔超平面分类器：

$$f(\boldsymbol{x}) = \text{sign}(\boldsymbol{w}^\top \phi(\boldsymbol{x}) + b)$$

其中：

- $\phi(\cdot)$ 为特征映射函数（本案例暂用线性核）
- \boldsymbol{w} 为法向量，决定分类边界方向
- b 为偏置项，控制边界位置

工程实现要点：

- 直接使用预处理后的特征矩阵，跳过了原始数据处理环节
- 实际工业场景需进行：

graph LR; A[原始数据] --> B[缺失值处理]; B --> C[特征标准化]; C --> D[维度约简/PCA]; D --> E[SVM建模]

模型定位与局限

支持向量机在本流程中承担**最终判别器**角色，其性能受限于：

1. 特征工程质量（如PCA/LDA的降维效果）
2. 核函数选择合理性（本案例为线性可分场景）
3. 正则化参数调节（控制模型复杂度与泛化能力的平衡）

注：本案例简化了特征工程环节，着重演示SVM的核心分类机制。