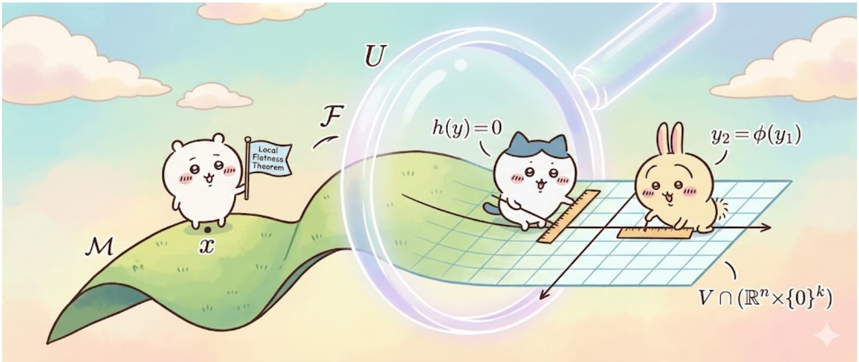
线性空间中的嵌入子流形

北极甜虾（南半球版）

2026.1.7

Manifold Optimization Notes



1 欧几里得空间

1.1 线性空间 (Linear Space)

在这里，我们说的线性空间是指实数域上的向量空间，我们总是会用 $\mathcal{E}$ 表示线性空间。常见的例子包括 $\mathbb{R}^d$、矩阵空间 $\mathbb{R}^{n\times p}$、对称矩阵空间 $\mathrm{Sym}(n)$ 等。

我们始终为 $\mathbb{R}^d$ 配备通常的拓扑。线性空间 $\mathcal{E}$ 通过基底与 $\mathbb{R}^d$ 的同构关系继承这一拓扑。这让我们能定义邻域、开集和连续性等概念。

1.2 欧几里得结构 (Euclidean Space)

流形优化的核心在于“长度”和“角度”的测量。这些概念由**内积（Inner Product）**定义。一个配备了内积 $\langle\cdot,\cdot\rangle$ 的线性空间称为欧几里得空间。我们定义以下标准内积：

- 对于向量 $u, v\in\mathbb{R}^d$：$\langle u,v\rangle = u^\top v = \sum_i u_i v_i$.

- 对于矩阵 $U, V \in \mathbb{R}^{n\times p}$: $\langle U, V\rangle = \mathrm{Tr}(U^\top V) = \sum_{i,j} U_{ij}V_{ij}$.

每一个内积都会诱导一个**范数 (Norm)**：$\|v\| = \sqrt{\langle v, v\rangle}$.

1.3 伴随算子与谱定理 (Adjoints and Spectral Theorem)

伴随算子与谱定理对理解 Hessian 矩阵至关重要。

- **伴随算子 $\mathcal{L}^*$：** 对于线性映射 $\mathcal{L} : \mathcal{E} \to \mathcal{F}$，其伴随映射满足 $\langle \mathcal{L}(u), v\rangle_{\mathcal{F}} = \langle u, \mathcal{L}^*(v)\rangle_{\mathcal{E}}$.

- **自伴随（对称）算子：** 如果 $\mathcal{A} = \mathcal{A}^*$，则称其为自伴随算子。

- **谱定理：** 欧几里得空间上的自伴随映射 $\mathcal{A}$ 拥有一组标准正交的特征向量基底，且特征值均为实数。如果所有特征值 $\lambda \geqslant 0$，则算子是半正定的。

线性空间上的范数会诱导线性映射的范数。我们定义线性映射 $\mathcal{L} : \mathcal{E} \to \mathcal{F}$ 的**算子范数 (Operator Norm)** 为

$$\|\mathcal{L}\| = \max_{u\in\mathcal{E}, u\neq 0} \frac{\|\mathcal{L}(u)\|_{\mathcal{F}}}{\|u\|_{\mathcal{E}}}.$$

- 等价地，$\|\mathcal{L}\|$ 是使得 $\|\mathcal{L}(u)\|_{\mathcal{F}} \leqslant \|\mathcal{L}\|\|u\|_{\mathcal{E}}$ 对所有 $u \in \mathcal{E}$ 都满足的最小的实数。

- $\mathcal{L}$ 的奇异值 (Singular Values) 为 $\mathcal{L}^* \circ \mathcal{L}$ 的特征值的非负平方根。$\|\mathcal{L}\|$ 也等于 $\mathcal{L}$ 奇异值中的最大值。

- 如果 $\mathcal{L}$ 是自伴随的，$\|\mathcal{L}\|$ 等于特征值中的最大值。

1.4 微分与梯度 (Differentials and Gradients)

我们先定义一些记号: 令 $F : A \to B$ 是一个函数，函数 F 在定义域的一个子集 $C \subseteq A$ 上的**限制 (Restriction)** 记为 $F|_C : C \to B$，其满足 $F|_C(x) = F(x), \forall x \in C$.

令 U, V 分别为两个线性空间 $\mathcal{E}, \mathcal{F}$ 的开子集，给定一个函数 $F : U \to V$，我们定义

- F 是**光滑的 (Smooth)** 或者 C^∞ 如果 F 在 U 上是无限次可微的。

- F 在 $x \in U$ 处是光滑的，如果存在一个邻域 $U' \ni x$ 使得 $F|_{U'}$ 是光滑的。

- F 是光滑的当且仅当在定义域内每一个点都是光滑的。

接下来我们给出微分，梯度，Hessian 的定义，这将是我们推广到流形上的定义的模板。

- **微分 (Differential)**：设 $F : U \to V$ 是光滑映射。在 x 处的微分是一个线性映射 $DF(x) : \mathcal{E} \to \mathcal{F}$，定义为

$$DF(x)[u] = \lim_{t \to 0} \frac{F(x + tu) - F(x)}{t}$$

- **欧几里得梯度 (Euclidean Gradient)**：对于光滑函数 $f : \mathcal{E} \to \mathbb{R}$，梯度 $\operatorname{grad} f(x)$ 是唯一的向量，满足对于所有方向 v，都有

$$\langle \operatorname{grad} f(x), v \rangle = Df(x)[v].$$

- **直观理解**：微分 $Df(x)[v]$ 是函数 f 在 x 点处往 v 方向上的变化率。梯度把这个“线性变化率”转化成空间中的一个具体向量。

- **欧几里得 Hessian：** 梯度的微分，即

$$\operatorname{Hess} f(x)[v] = D(\operatorname{grad} f)(x)[v] = \lim_{t \to 0} \frac{\operatorname{grad} f(x + tv) - \operatorname{grad} f(x)}{t}$$

  根据 Clairaut-Schwarz 定理，Hessian 是自伴随的：

$$\langle u, \operatorname{Hess} f(x)[v] \rangle = \langle \operatorname{Hess} f(x)[u], v \rangle.$$

定理 1: 伴随与转置, Exercise 3.9

假设 $\{u_1, \ldots, u_n\}$ 和 $\{v_1, \ldots, v_m\}$ 分别为两个欧几里得空间 $\mathcal{E}$ 和 $\mathcal{F}$ 的标准正交基。那么对于任意线性映射 $\mathcal{L} : \mathcal{E} \to \mathcal{F}$，以下结论成立：

1. 存在唯一的矩阵 $M \in \mathbb{R}^{m \times n}$ 满足

$$\mathcal{L}(u_i) = \sum_{j=1}^{m} M_{ji} v_j.$$

   我们称 M 是 $\mathcal{L}$ 的矩阵表示 (Matrix Representation)。

2. M 的显式表达式为

$$M_{ji} = \langle \mathcal{L}(u_i), v_j \rangle_{\mathcal{F}}\ , \forall j \in [m], i \in [n].$$

3. 伴随映射 $\mathcal{L}^*$ 的矩阵表示为 $M^\top$.

4. 一个线性映射 $\mathcal{A} : \mathcal{E} \to \mathcal{E}$ 是自伴随的当且仅当他的矩阵表示是对称的。

证明. 1. 由于 $\{v_1, \ldots, v_m\}$ 是 $\mathcal{F}$ 的一组基，根据基的定义，$\mathcal{F}$ 中的任何向量都可以唯一地表示为这组基向量的线性组

合。对于每一个 $\mathcal{L}(u_i) \in \mathcal{F}$（其中 $i = 1, \ldots, n$），存在唯一的标量 $M_{1i}, M_{2i}, \ldots, M_{mi}$ 使得：

$$\mathcal{L}(u_i) = \sum_{j=1}^{m} M_{ji} v_j$$

将这些标量排列成一个 $m \times n$ 的矩阵 M。将这些标量排列成一个 $m \times n$ 的矩阵 M，其中第 i 列包含了 $\mathcal{L}(u_i)$ 在基 $\{v_j\}$ 下的坐标。由于每个向量的坐标表示是唯一的，因此矩阵 M 也是唯一的。

2. 利用标准正交基的性质，我们可以得到 M_{ji} 的显式表达式：

$$\langle \mathcal{L}(u_i), v_j \rangle_{\mathcal{F}} = \left\langle \sum_{k=1}^{m} M_{ki} v_k, v_j \right\rangle_{\mathcal{F}} = \sum_{k=1}^{m} M_{ki} \langle v_k, v_j \rangle_{\mathcal{F}} = M_{ji}.$$

3. 伴随映射 $\mathcal{L}^* : \mathcal{F} \to \mathcal{E}$ 的定义是满足以下等式的唯一线性映射：

$$\langle \mathcal{L}(x), y \rangle_{\mathcal{F}} = \langle x, \mathcal{L}^*(y) \rangle_{\mathcal{E}}, \quad \forall x \in \mathcal{E}, y \in \mathcal{F}$$

设 N 是 $\mathcal{L}^*$ 关于基 $\{v_j\}$ 和 $\{u_i\}$ 的矩阵表示，即：

$$\mathcal{L}^*(v_j) = \sum_{k=1}^{n} N_{kj} u_k$$

根据 Part 2 中得到的表示法，矩阵 N 的元素 N_{ij}（第 i 行第 j 列）为：

$$N_{ij} = \langle \mathcal{L}^*(v_j), u_i \rangle_{\mathcal{E}} = \langle u_i, \mathcal{L}^*(v_j) \rangle_{\mathcal{E}}$$

利用伴随映射的定义：

$$N_{ij} = \langle \mathcal{L}(u_i), v_j \rangle_{\mathcal{F}}$$

再由 Part 2 的结论可知，$\langle \mathcal{L}(u_i), v_j \rangle_{\mathcal{F}} = M_{ji}$. 因此：

$$N_{ij} = M_{ji}$$

这说明 $N = M^\top$，即伴随映射的矩阵表示是原映射矩阵表示的转置。

4. 对于线性映射 $\mathcal{A} : \mathcal{E} \to \mathcal{E}$ 的情况下，我们在定义域和值域使用同一组标准正交基 $\{u_1, \ldots, u_n\}$。

   - $\Longrightarrow$：若 $\mathcal{A}$ 是自伴随的，则满足 $\mathcal{A} = \mathcal{A}^*$。设 $\mathcal{A}$ 的矩阵表示为 M。根据 Part 3，$\mathcal{A}^*$ 的矩阵表示为 $M^\top$。因为 $\mathcal{A} = \mathcal{A}^*$，它们的矩阵表示必然相等，即 $M = M^\top$。因此，M 是对称矩阵。
   - $\Longleftarrow$：若 $\mathcal{A}$ 的矩阵表示 M 是对称的，即 $M = M^\top$。由于 M 是 $\mathcal{A}$ 的矩阵表示，$M^\top$ 是 $\mathcal{A}^*$ 的矩阵表示，且矩阵表示与线性映射之间存在一一对应关系。由 $M = M^\top$ 可直接推出 $\mathcal{A} = \mathcal{A}^*$。因此，$\mathcal{A}$ 是自伴随的。

□

2 嵌入子流形

2.1 动机

给定线性空间 $\mathcal{E}$ 中的子集 $\mathcal{M}$. 当我们说 $\mathcal{M}$ 是"光滑的"，我们想表达 $\mathcal{M}$ 具有什么特征呢？

一个直觉是：**在每一点 $x \in \mathcal{M}$ 附近，可以被很好地线性近似**。假设 $\mathcal{M}$ 可以写为 $\mathcal{M} = \{x \in \mathcal{E} : h(x) = 0\}$，也就是说 $\mathcal{M}$ 由

函数 $h(x) = 0$ 确定。对于 x 附近的任意一点 $y = x + tv$，我们将 $h(y)$ 在 x 点进行一阶泰勒展开，我们得到

$$h(y) = h(x) + t \cdot Dh(x)[v] + r(t),$$

其中 $r(y)$ 是满足 $\lim_{t \to 0} \frac{\|r(t)\|}{t} = 0$ 的高阶余项。由于我们要寻找的是流形的“线性近似”，我们关注的是 $h(y)$ 的线性部分。令该线性近似等于 0 得到

$$h(x) + t \cdot Dh(x)[v] = 0.$$

考虑到 $x \in \mathcal{M}$，即 $h(x) = 0$，上述线性化方程简化为：

$$Dh(x)[v] = 0.$$

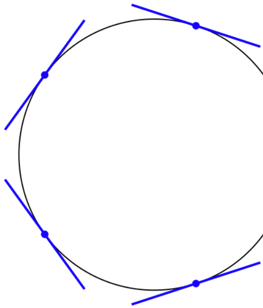
这个方程的解集为 $Dh(x)$ 的核（Kernel），其构成了 $\mathcal{M}$ 在 x 点处的**切空间（Tangent Space）**。

这一直觉告诉我们：如果一个集合 $\mathcal{M}$ 能由一个“足够好”的方程 $h(x) = 0$ 定义，那么我们就可以说 $\mathcal{M}$ 是光滑的，但是仅仅 h 是光滑的这一条性质并不足够。考虑以下例子：

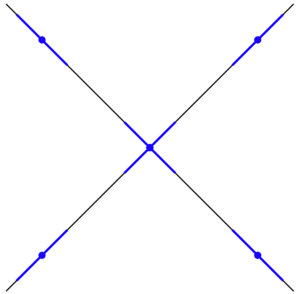
1. **Circle (圆):** $\mathcal{M} = \{x \in \mathbb{R}^2 : h(x) = x^\top x - 1 = 0\}$. 这是一个很好的流形。注意到微分为：

   $$Dh(x) = 2x.$$

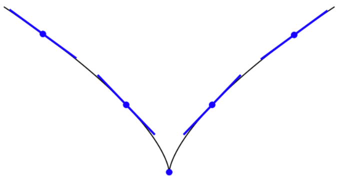
   其核为 $\ker Dh(x) = \{v \in \mathbb{R}^2 : x^\top v = 0\}$，即所有与位置向量 x 正交的方向。由于在圆上 $\|x\| = 1$，故 $Dh(x)$ 永远不会是零向量，即 $\operatorname{rank} Dh(x) = 1, \forall x \in \mathcal{M}$。这保证了圆在每一点处都有唯一的切线。



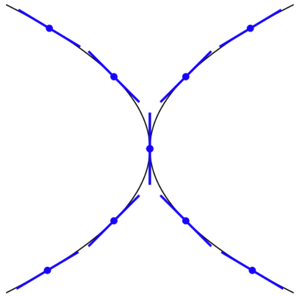
Circle: $x_1^2 + x_2^2 - 1 = 0$



Cross: $x_1^2 - x_2^2 = 0$



Cusp: $x_1^2 - x_2^3 = 0$



Double parabola: $x_1^2 - x_2^4 = 0$

2. **Cross (十字):** $\mathcal{M} = \{(x, y) \in \mathbb{R}^2 : h(x, y) = x^2 - y^2 = 0\}$. 这实际上是两条直线 $y = x$ 和 $y = -x$ 在原点相交。在原点 $(0, 0)$ 处计算微分：

$$Dh(x, y) = [2x, -2y] \implies Dh(0, 0) = [0, 0].$$

此时 $\ker Dh(0, 0) = \mathbb{R}^2$，$\operatorname{rank} Dh(0, 0) = 0$。由于秩发生塌缩（小于约束个数 $k = 1$），导致我们在交点处无法通过线性信息确定唯一的切线方向。

3. **Cusp (尖点)**：$\mathcal{M} = \{(x, y) \in \mathbb{R}^2 : h(x, y) = y^2 - x^3 = 0\}$. 在原点处 $Dh(0, 0) = [0, 0]$，计算得到 $\operatorname{rank} Dh(0, 0) = 0$. 虽然曲线连续，但“切线”在原点消失了。

4. **Double Parabolas (双抛物线):** $\mathcal{M} = \{(x, y) \in \mathbb{R}^2 : h(x, y) = y^2 - x^4 = 0\}$. 虽然两条抛物线在原点切向一致，但约束方程 h 在原点的微分 $Dh(0, 0) = [0, 0]$，导致 $\operatorname{rank} Dh(0, 0) = 0$。一阶导数无法捕捉到该点的几何走向。

这些例子告诉了我们：为了确保这个“切平面”真正存在且维度正确（即不会塌缩成一个点或扩张成整个空间），我们必须要求 $Dh(x)$ 是满秩的。

此外，嵌入子流形 $\mathcal{M}$ 通常难以由一个全局方程统一刻画。因此，定义中引入了邻域 U 的概念，并要求在该邻域内满足：$y \in \mathcal{M} \iff h(y) = 0$。这意味着在局部范围内，方程的零点集与流形是完全重合的。同时，描述同一子流形的局部定义函数 h 并不唯一，我们仅要求对于流形上的每一个点，这类函数在逻辑上是存在的。

2.2 嵌入子流形的定义

根据以上所有的讨论，我们可以给出线性空间中的嵌入子流形的定义。

定义 2: 线性空间中的嵌入子流形

设 $\mathcal{E}$ 是一个 d 维线性空间。一个非空子集 $\mathcal{M} \subseteq \mathcal{E}$ 被称为 $\mathcal{E}$ 的 n 维（光滑）嵌入子流形，如果其满足以下任一条件：

1. $n = d$ 且 $\mathcal{M}$ 是 $\mathcal{E}$ 中的一个开集。

2. $n = d - k$（$k \geqslant 1$），且对于 $\mathcal{M}$ 中的每一个点 x，都存在一个包含 x 的开邻域 $U \subseteq \mathcal{E}$ 以及一个光滑函数 $h : U \to \mathbb{R}^k$（称为局部定义函数），使得：

   - $U \cap \mathcal{M} = \{y \in U : h(y) = 0\}$
   - $\operatorname{rank} Dh(x) = k$ （即 $Dh(x)$ 是满秩的）

在接下来的内容中，我们将专注于线性空间中的嵌入子流形。相比于抽象流形，这类流形不仅几何直觉更加清晰，而且允许我们直接利用环境欧几里得空间的分析工具。为了表述简洁，后续我们将省略“光滑”二字，默认讨论的所有子流形及其关联函数均满足无限次可微性要求。

2.3 局部平坦化定理（Local Flatness Theorem）

既然一个嵌入子流形在局部可以表示为一个方程的解集，且定义这个方程的函数的微分是满秩的，那么其中 k 个变量就可以被其余 n 个变量唯一确定。直观理解就是流形在局部其实就是一段“平滑的函数曲线”，并且我们可以把这段曲线“按”到坐标轴上。为了保证微分性质，我们需要微分同胚的概念。

定义 3: 微分同胚

令 U, V 为两个线性空间上的开集，我们说函数 $F : U \to V$ 是一个微分同胚 (diffeomorphism)，如果 F 是双射，且 F 和 F^{-1} 都是光滑的。

定理 4: 局部平坦化定理（Local Flatness Theorem）

设 $\mathcal{E}$ 是维度为 d 的线性空间。一个子集 $\mathcal{M} \subseteq \mathcal{E}$ 是 $n = d - k$ 维嵌入子流形的充要条件是：对于每个 $x \in \mathcal{M}$，存在 x 的邻域 U, 开集 $V \subseteq \mathbb{R}^d$ 以及微分同胚 $F : U \to V$，使得：

$$F(\mathcal{M} \cap U) = V \cap (\mathbb{R}^n \times \{0\}^k).$$

证明. 根据嵌入子流形的定义，存在局部定义函数 $h : U_0 \to \mathbb{R}^k$，使得 $\mathcal{M} \cap U_0 = h^{-1}(0)$ 且 $\operatorname{rank} Dh(x) = k$。由于 $Dh(x) : \mathcal{E} \to \mathbb{R}^k$ 是满秩的线性映射，根据线性代数，我们可以通过选

择合适的基底，将 $\mathcal{E}$ 分解为直和：

$$\mathcal{E} = \mathcal{E}_1 \oplus \mathcal{E}_2 \cong \mathbb{R}^n \times \mathbb{R}^k$$

使得 $Dh(x)$ 在第二部分 $\mathcal{E}_2$ 上的限制 $Dh(x)|_{\mathcal{E}_2}$ 是一个可逆的 $k \times k$ 矩阵。

考虑方程 $h(y_1, y_2) = 0$，其中 $y_1 \in \mathbb{R}^n, y_2 \in \mathbb{R}^k$。因为 $h(x_1, x_2) = 0$ 且 $\frac{\partial h}{\partial y_2}(x_1, x_2)$ 可逆，隐函数定理指出：在 x 附近存在邻域 $U_1 \times U_2 \subseteq U_0$ 和一个光滑映射 $\phi : U_1 \to U_2$，使得在该邻域内：

$$h(y_1, y_2) = 0 \iff y_2 = \phi(y_1)$$

这意味着流形 $\mathcal{M}$ 在 x 附近局部上就是一个光滑函数的图像：

$$\mathcal{M} \cap (U_1 \times U_2) = \{(y_1, \phi(y_1)) : y_1 \in U_1\}$$

定义 $F : U_1 \times U_2 \to \mathbb{R}^n \times \mathbb{R}^k$ 如下：

$$F(y_1, y_2) = (y_1, y_2 - \phi(y_1)).$$

由于 ϕ 是光滑的，显然 F 也是光滑的。我们可以直接写出 F 的逆映射：

$$F^{-1}(z_1, z_2) = (z_1, z_2 + \phi(z_1))$$

该逆映射在相应的邻域上也是光滑的，因此 F 是一个微分同胚。

如果 $(y_1, y_2) \in \mathcal{M} \cap (U_1 \times U_2)$，因为 $y_2 = \phi(y_1)$, 代入 F 的公式得到

$$F(y_1, y_2) = (y_1, \phi(y_1) - \phi(y_1)) = (y_1, 0)$$

这恰好落在子空间 $\mathbb{R}^n \times \{0\}^k$ 中。反之，如果 $F(y_1, y_2) = (z_1, 0)$，则 $y_1 = z_1$ 且 $y_2 - \phi(y_1) = 0$，即 $y_2 = \phi(y_1)$，证明了该点一定在 $\mathcal{M}$ 上。 $\square$

参考文献

[Bou23] Nicolas Boumal. *An introduction to optimization on smooth manifolds*. Cambridge University Press, 2023.