一阶优化算法：最优性条件

北极甜虾 (南半球版)

2026 年 1 月 25 日

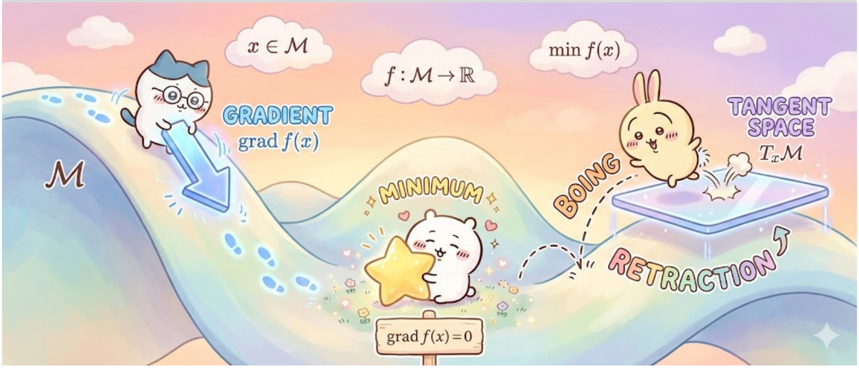
Manifold Optimization Notes

引言：流形上的无约束优化

流形上的一阶泰勒展开

临界点的定义

一阶最优性条件



1 引言：流形上的无约束优化

现在开始，我们将学习解决流形优化的核心任务：求解定义在流形 $\mathcal{M}$ 上的光滑函数 $f : \mathcal{M} \to \mathbb{R}$ 的优化问题：

$$\min_{x \in \mathcal{M}} f(x).$$

虽然这在欧几里得空间看来是一个带约束的优化问题，但在黎曼优化的视角下，我们可以将其视为流形上的**无约束优化**。

在理想情况下，我们希望寻找的是**全局极小值点 (Global Minimizer)**，即满足 $f(x) \leqslant f(y), \forall y \in \mathcal{M}$ 的点。然而，在非凸优化（流形优化通常是非凸的）中，找到全局最优通常是计算上不可行的。因此，我们退而求其次，寻找**局部极小值点 (Local Minimizer)**。

定义 1: 局部极小值点 (Local Minimizer)

点 $x \in \mathcal{M}$ 被称为 f 的一个**局部极小值点**，如果存在 x 的一个邻域 $U \subset \mathcal{M}$，使得对于所有 $y \in U$，都有 $f(x) \leqslant f(y)$。

拓扑的重要性 注意到局部极小值的定义仅依赖于流形的**拓扑结构**（即邻域的概念），而不需要黎曼度量。然而，为了设计寻找这些点的算法，我们需要引入几何和光滑结构。

为了分析算法产生的点序列 $x_0, x_1, x_2, \ldots$ 是否成功，我们需要严谨的收敛定义。

定义 2: 极限与聚点

设 $(x_k)_{k\in\mathbb{N}}$ 是流形 $\mathcal{M}$ 上的一个序列：

- **极限 (Limit)**：如果对于 x 的任意邻域 U，都存在 K 使得当 $k \geqslant K$ 时有 $x_k \in U$，则称序列收敛到 x，记为:

$$\lim_{k\to\infty} x_k = x \quad \text{或者} \quad x_k \to x.$$

  因为流形的拓扑结构是 Hausdorff 的，每一个收敛序列有且只有一个极限。

- **聚点 (Accumulation Point)**：若 x 的任意邻域都包含序列中无限多个点，则称 x 为该序列的一个聚点。

光滑结构的自由度 虽然“谁是极小值点”是由拓扑和函数本身决定的，但**黎曼度量是我们赋予流形的工具**。不同的度量会改变梯度的方向，从而改变算法的轨迹。我们可以根据问题的特性选择最合适的度量（如之前笔记中提到的 Fisher-Rao 度量），这正是黎曼优化的精妙之处。

2 流形上的一阶泰勒展开

在欧几里得空间中，泰勒展开是我们分析复杂函数在局部的性质的核心工具。但在流形 $\mathcal{M}$ 上，由于空间是弯曲的，我们不能直接进行 $x+v$ 这样的矢量相加。为了研究函数 f 在 x 附近的性质，我们必须借助一条经过 x 的光滑曲线，**将流形上的多变量问题限制在一条曲线上，转化成 $\mathbb{R}$ 到 $\mathbb{R}$ 的函数**。

定理 3: 一阶泰勒展开

令 $f:\mathcal{M}\to\mathbb{R}$ 是黎曼流形 $\mathcal{M}$ 上的光滑函数, $x\in\mathcal{M}$ 和切向量 $v\in T_x\mathcal{M}$. 对于任意光滑曲线 $c:I\to\mathcal{M}$ 满足 $c(0)=x$ 且 $c'(0)=v$，f 沿该曲线的一阶泰勒展开为：

$$f(c(t)) = f(x) + t\langle \operatorname{grad} f(x), v\rangle_x + O(t^2).$$

证明. 由于 $h(t)=f(c(t))$ 是从 $\mathbb{R}$ 到 $\mathbb{R}$ 的标准光滑函数，根据初等微积分中的泰勒公式：

$$h(t) = h(0) + th'(0) + \int_0^t (t-\tau)h''(\tau)\,d\tau.$$

根据黎曼梯度的定义：

$$h'(0) = \frac{d}{dt}f(c(t))\Big|_{t=0} = Df(x)[c'(0)] = \langle \operatorname{grad} f(x), v\rangle_x.$$

当 $t\to 0$ 时，余项部分为 $O(t^2)$，证毕。 □

推论 4: 流形上的中值定理

对于任意光滑曲线 $c:[0,1]\to\mathcal{M}$ 满足 $c(0)=x$ 且 $c(1)=y$，存在 $t\in(0,1)$ 使得

$$f(y) = f(x) + \langle \operatorname{grad} f(c(t)), c'(t)\rangle_{c(t)}.$$

直观理解：寻找最速下降方向

- **线性逼近：** 定理 3 告诉我们，在极小的步长 t 内，函数值的变化量几乎完全由内积 $\langle \operatorname{grad} f(x), v\rangle_x$ 决定。
- **下降方向：** 如果我们选择方向 $v = -\operatorname{grad} f(x)$，那么一阶项变为 $-t\|\operatorname{grad} f(x)\|_x^2$，这是局部下降最快的方向。这正是黎曼梯度下降的几何依据。

3 临界点的定义

直观上，如果一个点是局部的最优解，那么站在这个点向任何切向量方向看去，函数在这一瞬间都不应该有下降的趋势。

定义 5: 临界点 (Critical Point)

设 $f : \mathcal{M} \to \mathbb{R}$ 为光滑函数。一个点 $x \in \mathcal{M}$ 被称为 f 的一个**临界点**，如果它满足以下等价条件之一：

1. 对于任意满足 $c(0) = x$ 的光滑曲线 $c : I \to \mathcal{M}$，其在 $t = 0$ 处的一阶导数为零：
$$(f \circ c)'(0) = 0.$$
2. 对于所有的切向量 $v \in T_x\mathcal{M}$，函数在其方向上的微分（方向导数）为零：
$$Df(x)[v] = 0.$$

读者可自行证明两个条件等价。这里的第一个条件是从“动态”的角度看临界点：无论你沿着流形上的哪条路经过 x，在经过的一瞬间，高度的变化率都是 0。第二个条件则是从“静态”的切空间几何角度给出的描述。

4 一阶最优性条件

命题 6: 一阶最优性条件 (First-order Optimality Condition)

若 $x \in \mathcal{M}$ 是光滑函数 $f : \mathcal{M} \to \mathbb{R}$ 的局部极小值点，则 x 必为临界点。

证明. 设 x 是局部极小值点。对于任意 $v \in T_x\mathcal{M}$，取一条经过 x 且速度为 v 的光滑曲线 $c(t)$ （即 $c(0) = x, c'(0) = v$）。由于 x 是 f 在 $\mathcal{M}$ 上的局部极小值点，复合函数 $h(t) = f(c(t))$ 在 $t = 0$ 处必然取得了 $\mathbb{R}$ 上的局部极小值。根据欧几里得空间中的一阶最优性条件，必有 $h'(0) = 0$。即：

$$Df(x)[v] = h'(0) = 0.$$

由于此结论对任意 v 均成立，根据定义，x 是临界点。 □

虽然临界点可以通过微分 $Df(x)[v] = 0$ 来定义，但在实际算法设计中，我们更希望有一个像“梯度向量”一样的东西来指引方向。这就是黎曼梯度发挥作用的地方。

定理 7: 临界点与梯度的等价性

令 $\mathcal{M}$ 是一个黎曼流形。$x \in \mathcal{M}$ 是 $f : \mathcal{M} \to \mathbb{R}$ 的临界点，当且仅当在其黎曼度量下的梯度为零：

$$\operatorname{grad} f(x) = 0.$$

证明. 根据黎曼梯度的定义，对于所有 $v \in T_x\mathcal{M}$，有

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

($\Longrightarrow$) 若 x 是临界点，则对所有 v 有 $Df(x)[v] = 0$，故 $\langle \operatorname{grad} f(x), v \rangle_x = 0$. 取 $v = \operatorname{grad} f(x)$， 得 $\| \operatorname{grad} f(x)\|_x^2 = 0$， 即 $\operatorname{grad} f(x) = 0$.

($\Longleftarrow$) 若 $\operatorname{grad} f(x) = 0$，则内积显然为 0，故对所有 v 都有 $\mathrm{D}f(x)[v] = 0$. □

黎曼度量（内积）的**非退化性 (Non-degeneracy)** 决定了如果一个向量与空间中所有向量的内积都等于 0，那么这个向量本身必须为 0。因此定理 7 也说明了尽管梯度 $\operatorname{grad} f(x)$ 的具体数值取决于我们选择的黎曼度量，但一个点是否为临界点则是流形本身和函数 f 的固有属性。度量改变的是我们逼近临界点的“路径”，而不是临界点本身的位置。

参考文献

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