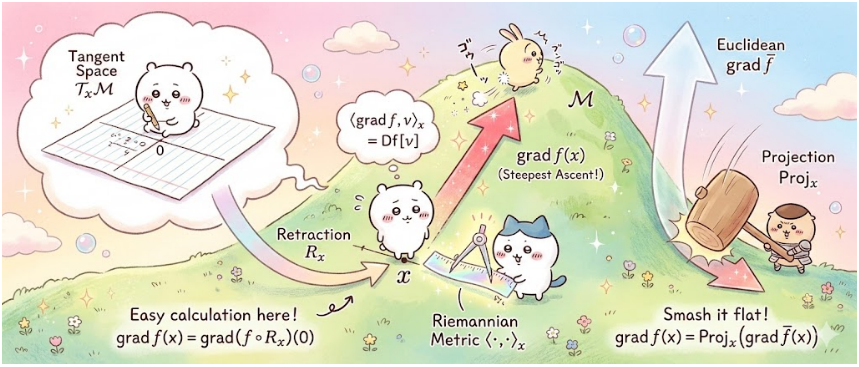
黎曼流形和黎曼梯度（一）

北极甜虾（南半球版）

2026.1.12

Manifold Optimization Notes



1 黎曼度量和黎曼流形

在之前的学习中，我们定义了标量场 f 的微分 $\mathrm{D}f(x)[v]$。它是一个线性算子，告诉我们：如果你沿着方向 v 走，函数值变化了多少。但在优化中，我们面临一个更核心的问题：在点 x 的切空间 $T_x\mathcal{M}$ 中，哪一个向量 v 是“最速下降方向”？

要回答这个问题，我们需要：

- **比较向量的大小：** 我们需要定义切向量的“长度”。
- **定义“正交”：** 我们需要知道哪个方向与等高线垂直。在欧几里得空间中，我们默认使用标准内积 $\langle u, v\rangle = u^\top v$。

注意在流形上，每一个切空间 $T_x\mathcal{M}$ 都是不同的，因此我们必须在每一个切空间上都指定一个内积，并且要求这个内积随点 x 的位置光滑地变化。

定义 1: 黎曼度量 (Riemannian metric)

流形 $\mathcal{M}$ 上的一个**黎曼度量**为每一个点 $x \in \mathcal{M}$ 指定了一个内积 $\langle\cdot,\cdot\rangle_x : T_x\mathcal{M} \times T_x\mathcal{M} \to \mathbb{R}$，且对于任意光滑向量场 $V, W \in \mathfrak{X}(\mathcal{M})$，标量函数 $x \mapsto \langle V(x), W(x)\rangle_x$ 都是光滑的。

每一个黎曼度量会诱导切空间上的范数 $\|v\|_x = \sqrt{\langle v, v\rangle_x}$.

定义 2: 黎曼流形 (Riemannian manifold)

一个**黎曼流形**是一个赋以黎曼度量的流形。

对于嵌入在欧几里得空间中的流形，我们可以通过最直观的方式获得这种度量：

定义 3: 诱导度量和黎曼子流形

对于嵌入子流形 $\mathcal{M} \subseteq \mathcal{E}$，最自然的选择是直接借用外在空间 $\mathcal{E}$ 的内积。如果我们定义 $\langle u, v\rangle_x = \langle u, v\rangle_{\mathcal{E}}$（即外在空间里的标准内积），则称 $\mathcal{M}$ 具有**诱导黎曼度量**。此时 $\mathcal{M}$ 称为 $\mathcal{E}$ 的一个**黎曼子流形**。

注意，这里“黎曼子流形”，并不单指带有任意黎曼结构的子流形，还限定了对度量的选择。

2 黎曼梯度 (Riemannian Gradient)

在之前的学习中，我们知道微分 $Df(x)$ 是一个线性算子，它描述了函数 f 在点 x 处的一阶近似。然而，在数值优化中，

我们需要一个具体的切向量来指示搜索方向。由于黎曼流形在每个点 x 的切空间 $T_x\mathcal{M}$ 上都定义了内积 $\langle\cdot,\cdot\rangle_x$，根据线性代数中的 Riesz 表示定理 (Riesz Representation Theorem)，对于流形上的每一个光滑向量场，它的黎曼梯度是存在且唯一的。

定理 4: Riesz 表示定理

令 H 是一个 Hilbert 空间，对于 H 上的任何连续线性泛函 $\phi \in H^*$，都存在唯一的向量 $g \in H$，使得对于所有 $v \in H$，都有 $\phi(v) = \langle g, v\rangle$.

定义 5: 黎曼梯度 (Riemannian Gradient)

令 $\mathcal{M}$ 是一个黎曼流形。对于一个光滑标量场 $f \in \mathfrak{F}(\mathcal{M})$，其在点 $x \in \mathcal{M}$ 处的**黎曼梯度** $\operatorname{grad} f(x)$ 是由以下等式唯一确定的切向量：

$$\langle \operatorname{grad} f(x), v\rangle_x = Df(x)[v], \quad \forall v \in T_x\mathcal{M}.$$

黎曼梯度之所以在优化中至关重要，是因为它在几何上代表了函数值增加最快的方向。

命题 6: 最速上升性质

在所有单位切向量 $v \in T_x\mathcal{M}$ 中，沿黎曼梯度方向 $v^* = \frac{\operatorname{grad} f(x)}{\|\operatorname{grad} f(x)\|_x}$ 的方向的微分 $Df(x)[v^*]$ 达到最大值, 即

$$\underset{v\in T_x\mathcal{M}:\|v\|_x=1}{\arg\max} \; Df(x)[v] = \frac{\operatorname{grad} f(x)}{\|\operatorname{grad} f(x)\|_x}.$$

证明. 根据梯度的定义和 Cauchy-Schwarz 不等式:

$$\mathrm{D}f(x)[v] = \langle \operatorname{grad} f(x), v \rangle_x \leqslant \| \operatorname{grad} f(x) \|_x \cdot \|v\|_x.$$

当 $\|v\|_x = 1$ 时，上式右侧最大值为 $\| \operatorname{grad} f(x)\|_x$。等号成立的充分必要条件是 v 与 $\operatorname{grad} f(x)$ 同向。 □

注解 7: 梯度取决于度量的选择

对于同一个函数 f，如果我们改变了流形上的黎曼度量 $\langle \cdot, \cdot \rangle_x$，计算出的梯度 $\operatorname{grad} f(x)$ 也会随之改变。在传统的欧几里得优化中，我们隐式地使用了标准内积。在黎曼优化中，通过精心选择度量（如自然梯度法中的 Fisher 信息度量），我们可以改变“最速”的定义，从而加速算法的收敛。

3 通过收缩映射进行梯度计算

在实际计算中，直接在流形上操作往往比较困难。收缩映射 (Retraction) 为我们提供了一个将流形上的函数转化为切空间（欧几里得空间）上函数的工具，从而简化梯度的计算。

命题 8

令 $\mathcal{M}$ 是一个黎曼流形，$f \in \mathfrak{F}(\mathcal{M})$ 为一个光滑标量场，R 为 $\mathcal{M}$ 上的收缩映射，那么

$$\operatorname{grad} f(x) = \operatorname{grad}(f \circ R_x)(0), \quad \forall x \in \mathcal{M}.$$

其中 $f \circ R_x : T_x\mathcal{M} \to \mathbb{R}$ 定义在欧几里得空间上，因此其梯度是标准意义下的梯度。

证明. 根据黎曼梯度的定义，$\operatorname{grad} f(x)$ 满足对于任意 $v \in T_x\mathcal{M}$:

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

由于 R_x 是一个收缩映射，它满足 $R_x(0) = x$ 且其在原点处的微分 $DR_x(0)$ 是恒等映射，即 $DR_x(0)[v] = v$。

利用复合函数求导的链式法则，考虑函数 $g = f \circ R_x$ 在 0 处沿方向 v 的微分：

$$\begin{aligned} Dg(0)[v] &= Df(R_x(0))[DR_x(0)[v]] \\ &= Df(x)[v]. \end{aligned}$$

在欧几里得空间 $T_x\mathcal{M}$ 中，经典梯度 $\operatorname{grad}(f \circ R_x)(0)$ 定义为满足以下条件的向量：

$$\langle \operatorname{grad}(f \circ R_x)(0), v \rangle_x = Dg(0)[v].$$

因此，对于所有 $v \in T_x\mathcal{M}$，均有

$$\langle \operatorname{grad}(f \circ R_x)(0), v \rangle_x = Df(x)[v] = \langle \operatorname{grad} f(x), v \rangle_x.$$

由内积的非退化性可知：

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

□

4 投影梯度公式

针对实际算法中最常见的嵌入子流形情形，我们只需将环境空间中的欧几里得梯度投影到切空间即可。

定理 9: 投影梯度公式

设 $\mathcal{M}$ 是 $\mathcal{E}$ 的黎曼子流形（使用诱导度量），$f \in \mathfrak{F}(\mathcal{M})$。若 $\bar{f}$ 是 f 在 $\mathcal{E}$ 的光滑延拓，则

$$\operatorname{grad} f(x) = \operatorname{Proj}_x(\operatorname{grad} \bar{f}(x))$$

其中 $\bar{f}:\mathcal{E}\to\mathbb{R}$ 定义在欧几里得空间上，因此其梯度是标准意义下的梯度。Proj_x 是到切空间 $T_x\mathcal{M}$ 的正交投影算子。

证明. 对于任意切向量 $v\in T_x\mathcal{M}$，根据梯度的定义以及 $\mathcal{M}$ 使用诱导度量的性质，有

$$\langle \operatorname{grad} f(x), v\rangle_{\mathcal{E}} = \mathrm{D}f(x)[v].$$

根据微分的延拓性质，对任意切向量 v，有 $\mathrm{D}f(x)[v]=\mathrm{D}\bar{f}(x)[v]$。而在欧几里得空间 $\mathcal{E}$ 中，微分与梯度的关系为

$$\mathrm{D}\bar{f}(x)[v] = \langle \operatorname{grad} \bar{f}(x), v\rangle_{\mathcal{E}}.$$

联立上述等式，我们得到

$$\langle \operatorname{grad} f(x), v\rangle_{\mathcal{E}} = \langle \operatorname{grad} \bar{f}(x), v\rangle_{\mathcal{E}}, \quad \forall v\in T_x\mathcal{M}.$$

利用正交投影算子 Proj_x 的性质（自伴随算子，且在切空间上是恒等映射），上式右侧可写为

$$\langle \operatorname{grad} \bar{f}(x), v\rangle_{\mathcal{E}} = \langle \operatorname{grad} \bar{f}(x), \mathrm{Proj}_x(v)\rangle_{\mathcal{E}} = \langle \mathrm{Proj}_x(\operatorname{grad} \bar{f}(x)), v\rangle_{\mathcal{E}}.$$

因此，对于所有 $v\in T_x\mathcal{M}$，都有：

$$\langle \operatorname{grad} f(x) - \mathrm{Proj}_x(\operatorname{grad} \bar{f}(x)), v\rangle_{\mathcal{E}} = 0.$$

由于 $\operatorname{grad} f(x)$ 和 $\mathrm{Proj}_x(\operatorname{grad} \bar{f}(x))$ 都属于切空间 $T_x\mathcal{M}$，根据内积的性质，两者必然相等：

$$\operatorname{grad} f(x) = \mathrm{Proj}_x(\operatorname{grad} \bar{f}(x)).$$

□

参考文献

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