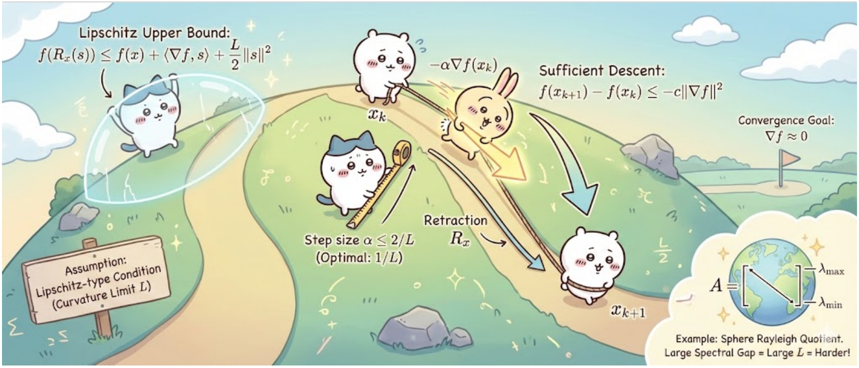
一阶优化算法: Lipschitz 型条件

北极甜虾 (南半球版)

2026 年 1 月 25 日

Manifold Optimization Notes

收敛性分析: Lipschitz 型条件 | **例子: 球面上的 Rayleigh 商**



1 收敛性分析：从充分下降到 Lipschitz 型条件

上一节我们证明黎曼梯度下降算法的收敛性的时候依赖于以下关键的条件：

假设 1: 充分下降条件

存在 $c > 0$ 使得序列 $\{x_k\}_{k\in\mathbb{N}}$ 满足：

$$f(x_{k+1}) - f(x_k) \leqslant -c\|\operatorname{grad} f(x_k)\|_{x_k}^2, \quad \forall k \in \mathbb{N}.$$

我们知道了只要满足**充分下降条件**，算法就会收敛, 为了在实际中选定步长 α 使得上述条件成立，我们必须限制目标函数沿流形移动时的“弯曲”程度。在欧几里得空间中，这由梯度的 Lipschitz 连续性保证, 在流形上，我们直接对**拉回映射 (Pullback)** $f \circ R$ 的增长进行限制。

假设 2: Lipschitz 型正则性条件

对于切丛 $T\mathcal{M}$ 的给定子集 S（通常包含算法产生的点的路径），存在常数 $L > 0$，使得对于所有 $(x, s) \in S$，满足：

$$f(R_x(s)) \leqslant f(x) + \langle \operatorname{grad} f(x), s \rangle_x + \frac{L}{2}\|s\|_x^2. \tag{1}$$

这个常数 L 扮演了“虚拟 Hessian 上界”的角色。它不仅取决于函数 f 本身的变化率，还捕捉了收缩映射 R_x 把直线“拉回”流形时产生的额外曲率。

命题 3: Lipschitz $\Longrightarrow$ 充分下降

若满足假设 2，且步长 $\alpha_k \in [\alpha_{\min}, \alpha_{\max}] \subset (0, 2/L)$ 范围内，则黎曼梯度下降产生的序列 $\{x_k\}_{k\in\mathbb{N}}$ 满足充分下降条件 (假设 1)。特别地，当取步长 $\alpha_k = 1/L$ $(\forall k \in \mathbb{N})$ 时，有

$$f(x_{k+1}) - f(x_k) \leqslant -\frac{1}{2L}\|\operatorname{grad} f(x_k)\|_{x_k}^2, \quad \forall k \in \mathbb{N}.$$

证明. 令公式 (1) 中的 $s_k = -\alpha_k \operatorname{grad} f(x_k)$，我们得到

$$\begin{aligned}
f(x_{k+1}) &= f(R_{x_k}(s_k)) \\
&\leqslant f(x_k) + \langle \operatorname{grad} f(x_k), -\alpha_k \operatorname{grad} f(x_k) \rangle_{x_k} + \frac{L}{2}\|\operatorname{grad} f(x_k)\|_{x_k}^2 \\
&\leqslant f(x_k) - \alpha_k\|\operatorname{grad} f(x_k)\|_{x_k}^2 + \frac{L\alpha_k^2}{2}\|\operatorname{grad} f(x_k)\|_{x_k}^2.
\end{aligned}$$

整理后得到

$$f(x_{k+1}) - f(x_k) \leqslant -\left(\alpha_k - \frac{L\alpha_k^2}{2}\right)\|\operatorname{grad} f(x_k)\|_{x_k}^2.$$

注意到 $\alpha_k - \frac{L\alpha_k^2}{2}$ 是个二次函数，其最小值在区间端点处取到，因此存在 $c = \min\{\alpha_{\min} - \frac{L}{2}\alpha_{\min}^2, \alpha_{\max} - \frac{L}{2}\alpha_{\max}^2\}$，使得假设 1 满足。 □

在欧几里得空间中，只要 ∇f 是 L-Lipschitz 连续的，类似命题 3 的下降引理就会成立。在流形上，由于 R 的存在，这一条件不仅取决于 f 的光滑性，还取决于收缩映射 R 的二阶导数（即它把直线"弯"回流形的剧烈程度）。公式 (1) 本质上是给出了一个二次上界，保证了只要步长足够小，线性项的下降（$-\alpha\|\operatorname{grad} f\|^2$）总能抵消掉二次项的上升（$+\frac{L\alpha^2}{2}\|\operatorname{grad} f\|^2$）。

定理 4: 基于 Lipschitz 型正则性条件的收敛定理

如果下列条件满足：

- 存在 $f_{\text{low}} \in \mathbb{R}$ 使得 $f(x) \geqslant f_{\text{low}}, \forall x \in \mathcal{M}$.
- Lipschitz 型正则性条件 (假设 2)
- 步长设置为 $\alpha_k = 1/L, \forall k \in \mathbb{N}$

那么对于任意整数 $K \geqslant 1$，黎曼梯度下降产生的序列满足

$$\min_{0\leqslant k\leqslant K-1} \|\operatorname{grad} f(x_k)\|_{x_k} \leqslant \sqrt{\frac{2L(f(x_0) - f_{\text{low}})}{K}}.$$

证明. 该定理由命题 3 以及上一节笔记中的定理 3 直接得到。 □

2 例子：球面上的 Rayleigh 商

考虑球面 S^{n-1} 上的函数 $f(x) = \frac{1}{2}x^\top Ax$（其中 A 为对称矩阵），配备收缩映射 $R_x(s) = \frac{x+s}{\|x+s\|}$。我们来确定一个全局的 L。

命题 5: 球面 Rayleigh 商的 Lipschitz 常数, Exercise 4.11

对于球面上的 Rayleigh 商，取常数

$$L = \lambda_{\max}(A) - \lambda_{\min}(A)$$

能使假设 2 在整个切丛上成立。

证明. 由 $x \in S^{n-1}$ 知 $\|x\| = 1$。对于切向量 $s \in T_x S^{n-1}$，有 $x^\top s = 0$. 收缩映射可写为

$$R_x(s) = \frac{(x+s)}{\sqrt{1+\|s\|^2}}.$$

代入得

$$f(R_x(s)) = \frac{1}{2}\frac{(x+s)^\top A(x+s)}{1+\|s\|^2}.$$

球面上的黎曼梯度为

$$\operatorname{grad} f(x) = Ax - (x^\top Ax)x,$$

因为 $s \perp x$，所以 $f(R_x(s))$ 的一阶项为

$$\langle \operatorname{grad} f(x), s\rangle_x = s^\top Ax.$$

我们要寻找 L 使得

$$\frac{1}{2}\frac{x^\top Ax + 2s^\top Ax + s^\top As}{1+\|s\|^2} \leqslant \frac{1}{2}x^\top Ax + s^\top Ax + \frac{L}{2}\|s\|^2.$$

令 $t = \|s\|, v = s/t$（v 是单位切向量），整理得到

$$x^\top Ax + 2tv^\top Ax + t^2 v^\top Av \leqslant (x^\top Ax + 2tv^\top Ax + Lt^2)(1+t^2)$$

化简后，要求对所有 $t \geqslant 0$ 满足

$$Lt^4 + 2(v^\top Ax)t^3 + (L + x^\top Ax - v^\top Av)t^2 \geqslant 0.$$

由于 $t^2 > 0$，这等价于二次函数

$$Q(t) = Lt^2 + 2(v^\top Ax)t + (L + x^\top Ax - v^\top Av) \geqslant 0.$$

该二次式非负的充要条件是判别式 $\Delta \leqslant 0$（且 $L > 0$）：

$$\begin{aligned}&(v^\top Ax)^2 - L(L + x^\top Ax - v^\top Av) \leqslant 0\\ \Longrightarrow\ & L^2 + L(x^\top Ax - v^\top Av) - (v^\top Ax)^2 \geqslant 0.\end{aligned}$$

记 $\Delta\lambda = \lambda_{\max}(A) - \lambda_{\min}(A)$．通过对特征值进行平移（不改变梯度和 L），设 $\lambda_{\min} = 0, \lambda_{\max} = \Delta\lambda$。此时 $x^\top Ax \in [0, \Delta\lambda]$，$v^\top Av \in [0, \Delta\lambda]$，且

$$(v^\top Ax)^2 \leqslant (x^\top Ax)(\Delta\lambda - x^\top Ax).$$

取 $L = \Delta\lambda$，代入验证发现不等式总是成立。 $\square$

这个结果很有意思，注意到在这个问题中谱间距决定了优化的难度。常数 L 直接由 A 的特征值范围决定。如果矩阵非常“扁”（条件数大），L 就大，意味着我们需要更小的步长来保证下降。

参考文献

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