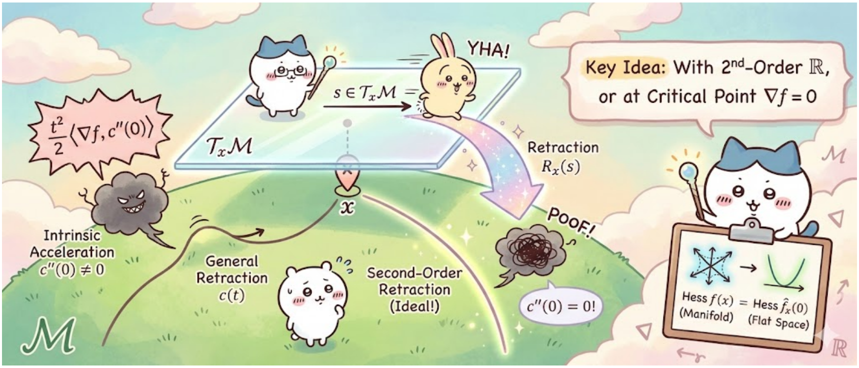
二阶收缩映射

北极甜虾 (南半球版)

2026 年 1 月 28 日



1 二阶收缩映射的动机与定义

在流形优化中，我们用**收缩映射（Retraction）**R 将切空间中的步长 $s \in T_x\mathcal{M}$ 映射回流形上：

$$x^+ = R_x(s).$$

因此，理解**拉回映射（Pullback）**

$$\hat{f}_x : T_x\mathcal{M} \to \mathbb{R}, \qquad \hat{f}_x(s) = (f \circ R_x)(s)$$

在 $s = 0$ 附近的二阶结构，是构造/分析二阶算法，比如 Newton、Trust-region 等的关键。

令 $v \in T_x\mathcal{M}$，考虑**收缩曲线 (Retraction Curve)**

$$c(t) = R_x(tv), \qquad c(0) = x, \;\; c'(0) = v.$$

由上一节沿曲线的二阶泰勒展开式直接得到：

$$f(R_x(tv)) = f(x) + t\langle \operatorname{grad} f(x), v\rangle_x + \frac{t^2}{2}\langle \operatorname{Hess} f(x)[v], v\rangle_x + \frac{t^2}{2}\langle \operatorname{grad} f(x), c''(0)\rangle_x + O(t^3).$$

其中 $c''(0) = \frac{D}{dt}c'(0)$ 是曲线 c 在 $t = 0$ 的**内蕴加速度 (Intrinsic Acceleration)**。

注意最后一项 $\frac{t^2}{2}\langle \operatorname{grad} f(x), c''(0)\rangle_x$ 是不希望出现的，因为它依赖于所选收缩映射 R，而不是仅依赖于流形几何与 f 的二阶信息）。

该项在两类情形会消失：

$$\text{(i)}\ \operatorname{grad} f(x) = 0 \quad \text{或} \quad \text{(ii)}\ c''(0) = 0.$$

前者对应“临界点附近”的分析，后者提示我们可以对收缩映射提出更强的几何要求，从而在任意点消除该项。

定义 1: 二阶收缩映射（second-order retraction）

令 $\mathcal{M}$ 是黎曼流形。称收缩映射 R 为**二阶收缩映射（second-order retraction）**，若对任意 $x \in \mathcal{M}$ 与任意 $v \in T_x\mathcal{M}$，收缩曲线

$$c(t) = R_x(tv)$$

在 $t = 0$ 处满足的内蕴加速度为 0：

$$c''(0) = 0.$$

2 一些例子和性质

例子 2: 球面上的归一化收缩映射是二阶的

在单位球面 S^{d-1} 上考虑常用收缩映射

$$R_x(v) = \frac{x+v}{\|x+v\|}.$$

令 $c(t) = R_x(tv)$。计算可得其（欧几里得）二阶导满足

$$\ddot{c}(0) = -\|v\|^2 x,$$

它与 T_xS^{d-1} 正交。因此内蕴加速度是切向投影：

$$c''(0) = \mathrm{Proj}_x(\ddot{c}(0)) = \mathrm{Proj}_x(-\|v\|^2 x) = 0.$$

故该收缩映射是二阶收缩映射。

命题 3: 两种重要特例

设 $\mathcal{M}$ 是黎曼流形，R 是任意收缩映射，$f: \mathcal{M} \to \mathbb{R}$ 光滑，$\widehat{f}_x = f \circ R_x$.

(1) **临界点处（任意收缩映射都能用）**：若 x 是临界点（$\mathrm{grad}\, f(x) = 0$），则

$$f(R_x(s)) = f(x) + \frac{1}{2}\langle \mathrm{Hess}\, f(x)[s], s\rangle_x + O(\|s\|_x^3).$$

(2) **二阶收缩映射**: 若 R 是二阶收缩映射，则对任意 $x \in \mathcal{M}$ 都有

$$f(R_x(s)) = f(x) + \langle \mathrm{grad}\, f(x), s\rangle_x + \frac{1}{2}\langle \mathrm{Hess}\, f(x)[s], s\rangle_x + O(\|s\|_x^3).$$

证明. 将上一节的收缩曲线展开式写成 $s = tv$ 的形式即可。

- 若 $\operatorname{grad} f(x) = 0$，则加速度项自动消失，得到结论 (1)。
- 若 R 二阶，则 $c''(0) = 0$，从而加速度项消失，得到结论 (2)。

并将 $O(t^3)$ 重写为 $O(\|s\|_x^3)$（因 $s = tv$ 且局部等价）。 □

命题 3 说明**在临界点附近**，即便只用一阶收缩映射，二阶展开也不会被加速度项影响；而若我们拥有**二阶收缩映射**，则**在任意点**都可以直接把 $\hat{f}_x$ 当作外在空间的函数来做标准二阶近似。

命题 4: 用 pullback 在切空间中计算黎曼 Hessian

若 R 是二阶收缩映射，或 x 是临界点（$\operatorname{grad} f(x) = 0$），则

$$\operatorname{Hess} f(x) = \operatorname{Hess}(f \circ R_x)(0),$$

右侧是定义在欧式空间 $T_x\mathcal{M}$ 上的**经典 Hessian**。

证明. 记 $\hat{f}_x(s) = f(R_x(s))$。

- 若 R 二阶，由命题 3(2)：

$$\hat{f}_x(s) = f(x) + \langle \operatorname{grad} f(x), s\rangle_x + \frac{1}{2}\langle \operatorname{Hess} f(x)[s], s\rangle_x + O(\|s\|_x^3).$$

在欧式空间 $T_x\mathcal{M}$ 对 s 求导（并用 $\operatorname{Hess} f(x)$ 自伴随性）：

$$\operatorname{grad} \hat{f}_x(s) = \operatorname{grad} f(x) + \operatorname{Hess} f(x)[s] + O(\|s\|_x^2),$$

$$\operatorname{Hess} \hat{f}_x(s)[\dot{s}] = \operatorname{Hess} f(x)[\dot{s}] + O(\|s\|_x\|\dot{s}\|_x).$$

令 $s = 0$，得 $\operatorname{Hess} \hat{f}_x(0) = \operatorname{Hess} f(x)$。

- 若 $\operatorname{grad} f(x) = 0$，同理由命题 3(1) 出发即可。

$\square$

注解 5: 临界点处 Hessian 对度量的不变性

临界点与(一阶)收缩映射的定义不依赖于黎曼度量。命题 4 进一步表明：在临界点 x 处，$\operatorname{Hess} f(x)$ 对所选黎曼度量的依赖只通过切空间上的内积 $\langle \cdot, \cdot \rangle_x$ 体现。特别地，$\operatorname{Hess} f(x)$ 的符号结构（signature）（正/零/负特征值的个数）与我们选取的黎曼度量无关。

注解 6: 二阶收缩映射的加速度保持性质, Exercise 5.46

固定 $x \in \mathcal{M}$，取切空间曲线 $w : I \to T_x\mathcal{M}$，满足 $w(0) = 0$，并诱导流形曲线

$$c(t) = R_x(w(t)).$$

显然有 $c(0) = x$ 且 $c'(0) = w'(0)$。若 R 为二阶收缩映射，则还有

$$c''(0) = w''(0).$$

一个直接推论：给定任意 $u, v \in T_x\mathcal{M}$，可以构造曲线

$$c(t) = R_x\left(tu + \frac{t^2}{2}v\right)$$

使得 $c'(0) = u$ 且 $c''(0) = v$。这从侧面解释了二阶收缩映射为何能对齐切空间中的二阶信息。

参考文献

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