

# 流形优化的一些例子（二）

北极甜虾（南半球版）

2026.1.6

Manifold Optimization Notes

旋转同步

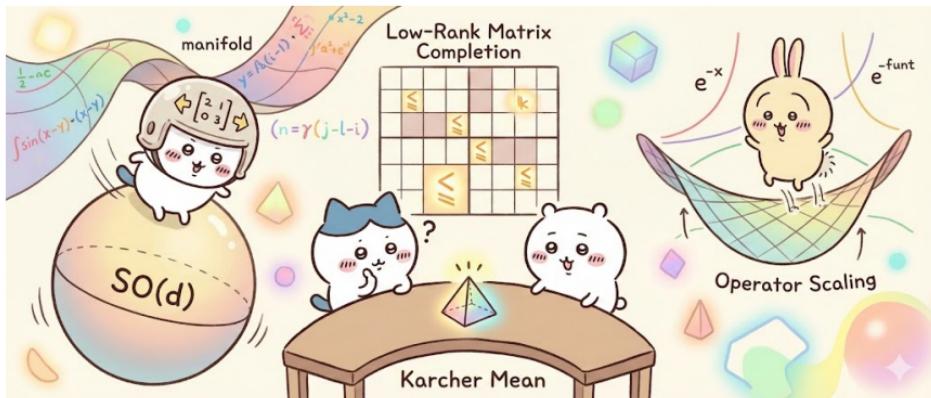
低秩矩阵补全

高斯混合模型

光滑半正定规划

算子缩放

Karcher 均值



## 1 旋转同步与特殊正交群

### 问题 1: 旋转同步 (Synchronization of Rotations)

假设有  $n$  个观测对象，其未知的绝对旋转状态为  $R_1, \dots, R_n \in SO(d)$ 。我们能够观测到部分对象两两之间  $(i, j) \in G$  的相对旋转测量值  $H_{ij} \in SO(d)$ ，其中

$$H_{ij} \approx R_i R_j^{-1}.$$

我们的目标是根据这些带噪声的相对测量值来估计所有对象的绝对旋转状态  $R_1, \dots, R_n$ .

**特殊正交群 (Special Orthogonal Group)** 数学上, 旋转通过特殊正交群  $\text{SO}(d)$  来描述, 其定义为:

$$\text{SO}(d) = \{R \in \mathbb{R}^{d \times d} : R^\top R = I_d \text{ 且 } \det(R) = 1\}.$$

**目标函数的构建** 为了衡量估计值  $\hat{R}_i$  与测量值  $H_{ij}$  的拟合程度, 我们利用 Frobenius 范数构建最小二乘问题。注意到理想情况下  $\hat{R}_i = H_{ij}\hat{R}_j$ , 因此目标函数设定为

$$\min_{\hat{R}_1, \dots, \hat{R}_n \in \text{SO}(d)} \sum_{(i,j) \in G} \|\hat{R}_i - H_{ij}\hat{R}_j\|_F^2.$$

该模型也常用于同步定位与建图 (SLAM) 中, 帮助机器人校正姿态估计的漂移。

## 流形优化的视角

- **搜索空间:** 该问题的决策变量是  $n$  个旋转矩阵的组合, 其搜索空间是  $n$  个特殊正交群的笛卡尔积, 记作  $(\text{SO}(d))^n$ 。
- **流形性质:**  $\text{SO}(d)$  是  $\mathbb{R}^{d \times d}$  的一个嵌入子流形, 同时它也是一个李群 (Lie group)。由于流形乘积仍为流形, 因此  $(\text{SO}(d))^n$  是一个光滑的, 维数为  $n \cdot \frac{d(d-1)}{2}$  的流形。

## 2 低秩矩阵补全: 固定秩流形

### 问题 2: 低秩矩阵补全 (Low-rank Matrix Completion)

一个矩阵  $M \in \mathbb{R}^{m \times n}$  的大部分项是不可观测的。令  $\Omega$  为已知项的指标的集合。基于数据具有低秩结构的假设, 目标是通过最小化观测位置的误差平方和来预测缺失的项。

**低秩模型与优化问题** 如果矩阵  $M$  的秩为  $r$ , 则它可以表示为  $M = LR^\top$ , 其中  $L \in \mathbb{R}^{m \times r}$  和  $R \in \mathbb{R}^{n \times r}$  是满秩因子矩阵。在实际应用中, 我们将问题建模为以下带秩约束的最小二乘问题:

$$\min_{X \in \mathbb{R}^{m \times n}} \sum_{(i,j) \in \Omega} (X_{ij} - M_{ij})^2 \quad \text{s.t.} \quad \text{rank}(X) = r.$$

## 流形优化的视角

- **搜索空间:** 该问题的可行域是所有大小为  $m \times n$  且秩正好为  $r$  的矩阵集合, 记作  $\mathbb{R}_r^{m \times n}$ , 该空间是  $\mathbb{R}^{m \times n}$  的一个嵌入子流形, 这一点由隐函数定理可知。
- **重要提醒:** 需要注意的是, 秩不超过  $r$  的矩阵集合 (即  $\mathbb{R}_r^{m \times n}$  的闭包) 并不是一个流形, 因为它在秩降低的点处存在奇异性。例如考虑  $2 \times 2$  矩阵中秩  $\leq 1$  的集合, 这由方程  $ad - bc = 0$  定义。在  $\mathbb{R}^4 (\simeq \mathbb{R}^{2 \times 2})$  中, 这是一个二次锥面。在零矩阵 (秩为 0) 处, 它是锥的顶点, 我们无法定义一个唯一的切平面来逼近这个点。

**商流形视角** 由于分解  $X = LR^\top$  具有不变性对于任何秩为  $r$  的可逆矩阵  $J \in GL(r)$ , 如果令  $\tilde{L} = LJ^{-1}$  且  $\tilde{R} = RJ^\top$ , 那么:

$$\tilde{L}\tilde{R}^\top = (LJ^{-1})(RJ^\top)^\top = LJ^{-1}JR^\top = LR^\top = X$$

我们定义  $(L, R) \sim (L', R')$  当且仅当它们对应同一个  $X$ 。我们可以在这个等价关系诱导的商空间中进行优化。

### 3 高斯混合模型与正定矩阵

#### 问题 3: 高斯混合模型 (Gaussian Mixture Models)

假设数据  $x_1, \dots, x_n \in \mathbb{R}^d$  是独立地从  $K$  个高斯分布的混合中采样得到的。每个数据点的概率密度函数具有如下形式：

$$f(x) = \sum_{k=1}^K w_k \frac{1}{\sqrt{(2\pi)^d \det(\Sigma_k)}} \exp\left(-\frac{1}{2}(x - \mu_k)^\top \Sigma_k^{-1} (x - \mu_k)\right)$$

其中  $\mu_k \in \mathbb{R}^d$  是均值,  $w_k$  是混合权重,  $\Sigma_k$  是协方差矩阵。目标是根据观测值估计这些参数。

**参数约束与搜索空间** 在估计过程中, 参数必须满足特定的几何约束:

- **对称正定约束:** 协方差矩阵  $\Sigma_k$  必须属于对称正定矩阵集合:

$$\text{Sym}(d)_+ = \{X \in \mathbb{R}^{d \times d} : X = X^\top \text{ 且 } X \succ 0\}.$$

- **单纯形约束:** 混合权重  $w$  必须属于正单纯形 (Positive Simplex):

$$\Delta_{K-1}^+ = \{w \in \mathbb{R}^K : w_k > 0 \text{ 且 } \sum_{k=1}^K w_k = 1\}.$$

**最大似然估计 (MLE)** 对于给定观测集  $\{x_i\}_{i=1}^n$ , 参数的估计可以通过最大化对数似然函数来实现:

$$\max_{\substack{\mu_1, \dots, \mu_K \in \mathbb{R}^d \\ \Sigma_1, \dots, \Sigma_K \in \text{Sym}(d)_+ \\ w \in \Delta_{K-1}^+}} \sum_{i=1}^n \log \left( \sum_{k=1}^K w_k \frac{1}{\sqrt{(2\pi)^d \det(\Sigma_k)}} \exp\left(-\frac{1}{2}(x_i - \mu_k)^\top \Sigma_k^{-1} (x_i - \mu_k)\right) \right). \quad (1)$$

## 流形优化的视角

- **流形构造:** 问题的搜索空间  $\mathbb{R}^{d \times K} \times (\text{Sym}(d)_+)^K \times \Delta_{K-1}^+$  可以被赋予流形结构, 因为  $\text{Sym}(d)_+$  和  $\Delta_{K-1}^+$  都是光滑流形。
- **计算考量:** 虽然直接求解公式 (1) 较为困难, 但在流形上进行重新表述 (例如利用某些对称性) 可以获得更好的计算性能。

## 4 光滑半正定规划

### 问题 4: 半正定规划 (SDP)

标准的半正定规划问题定义为:

$$\begin{aligned} & \min_{X \in \text{Sym}(n)} \quad \langle C, X \rangle \\ & \text{s.t.} \quad \mathcal{A}(X) = b, \quad X \succeq 0. \end{aligned}$$

其中  $\mathcal{A}(X) = [\langle A_1, X \rangle, \dots, \langle A_m, X \rangle]^\top$  是线性映射. 当  $n$  较大时, 处理  $X \succeq 0$  约束和  $O(n^2)$  级别的变量具有极大的计算挑战.

**Burer-Monteiro 方法** 为了降低维度并消除硬约束, 可以将  $X$  分解为  $X = YY^\top$ , 其中  $Y \in \mathbb{R}^{n \times p}$ . 此时问题转化为:

$$\begin{aligned} & \min_{Y \in \mathbb{R}^{n \times p}} \quad \langle CY, Y \rangle \\ & \text{s.t.} \quad \mathcal{A}(YY^\top) = b. \end{aligned}$$

根据 **Pataki-Barvinok 界**, 若可行域紧致, 只要  $p(p+1)/2 \geq m$ , 该低秩形式就与原 SDP 等价.

**流形优化的视角** 从几何角度看，该问题的搜索空间是满足线性约束的矩阵集合：

- **搜索空间：**  $\mathcal{M} = \{Y \in \mathbb{R}^{n \times p} : \mathcal{A}(YY^\top) = b\}$ .
- **流形性质：** 若映射  $Y \mapsto \mathcal{A}(YY^\top)$  的微分在  $\mathcal{M}$  上处处满秩，则  $\mathcal{M}$  是  $\mathbb{R}^{n \times p}$  的一个嵌入子流形.

**良态非凸性 (Benign Non-convexity)** 尽管分解引入了非凸性，但在特定条件下（如  $\mathcal{M}$  紧致且  $p$  足够大），对于泛函代价矩阵  $C$ ，该流形优化问题不存在伪局部最小值. 任何满足一阶和二阶必要条件的点均是全局最优解. 这使得流形优化算法（如 Riemannian 信赖域法）求解大规模 SDP 变得非常高效.

## 5 算子缩放：测地凸性

### 问题 5: 算子缩放 (Operator Scaling)

给定矩阵  $A_1, \dots, A_m \in \mathbb{C}^{n \times n}$ . 我们的目标是找到正定矩阵  $X$  和  $Y$ ，使得缩放后的算子  $\tilde{A}_i = Y^{1/2} A_i X^{1/2}$  是双随机的 (Doubly Stochastic)，即满足：

$$\sum_{i=1}^m \tilde{A}_i \tilde{A}_i^\dagger = I_n \quad \text{且} \quad \sum_{i=1}^m \tilde{A}_i^\dagger \tilde{A}_i = I_n.$$

**容量 (Capacity) 与目标函数** 定义的正算子 (Positive Operator)  $T(X) = \sum_{i=1}^m A_i X A_i^\dagger$ . 算子  $T$  的容量定义为以下优化问题的下确界：

$$cap(T) = \inf_{X > 0, \det(X)=1} \det(T(X)).$$

如果  $\text{cap}(T)$  在  $X^*$  处取得最小值，那么令  $Y^* = T(X^*)^{-1}$  就会使得缩放后的算子  $\tilde{A}_i = (Y^*)^{1/2} A_i (X^*)^{1/2}$ . 我们也可以将其等价地表示为最小化对数容量函数：

$$\min_{X \in \text{Sym}(n)_+} f(X) = \log \det \left( \sum_{i=1}^m A_i X A_i^\dagger \right) - \log \det(X).$$

## 流形优化的视角

- **非凸性与测地凸性：**在传统的欧几里得几何下， $f(X)$  通常是非凸的。然而，在正定矩阵流形  $\text{Sym}(n)_+$  配备自然双曲度量（黎曼度量）时， $f(X)$  是 **测地凸 (Geodesically Convex)** 的。
- **二阶算法的指数级提升：**相比于传统的 Sinkhorn 类算法（一阶方法，误差依赖于  $1/\epsilon$ ），测地二阶方法可以实现对数级误差依赖  $\text{poly}(\log(1/\epsilon))$  [AZGL<sup>+</sup>18]。

## 6 Karcher Mean：正定矩阵流形的几何平均

### 问题 6: 黎曼质心 (Riemannian Center of Mass)

给定一组对称正定 (SPD) 矩阵  $A_1, \dots, A_K \in \text{Sym}(n)_{++}$ ，我们的目标是找到一个“平均”矩阵  $\mu$ ，使其在某种度量下到所有观测点的距离平方和最小。

**SPD 流形与仿射不变度量** 在 SPD 矩阵空间上，Frobenius 范数会导致平均后的矩阵行列式显著增大。为了保持逆不变性等优良性质，我们赋予流形 **仿射不变度量 (Affine-Invariant Metric)**：

$$g_X(\xi_X, \eta_X) = \text{trace}(\xi_X X^{-1} \eta_X X^{-1}).$$

该度量诱导出的两点  $p, q$  间的 测地距离 为：

$$\delta(p, q) = \|\log(p^{-1/2} q p^{-1/2})\|_F.$$

**目标函数的构建** Karcher 均值被定义为以下优化问题的全局唯一最小值点：

$$\min_{X \in \text{Sym}(n)_{++}} F(X) = \frac{1}{2K} \sum_{i=1}^K \delta^2(X, A_i).$$

## 流形优化的视角

- **测地凸性：**在仿射不变度量下，函数  $F(X)$  具有强测地凸性。这意味着流形优化算法能够避免陷入局部最优，具有与欧几里得凸优化类似的收敛保证。
- **Hessian 良态性：**研究表明，Riemannian Hessian 在该流形上的条件数通常非常稳定，且不会像 Euclidean Hessian 那样随矩阵条件数恶化而急剧坍塌 [YHAG20]。

## 参考文献

- 
- [AZGL<sup>+</sup>18] Zeyuan Allen-Zhu, Ankit Garg, Yuanzhi Li, Rafael Oliveira, and Avi Wigderson. Operator scaling via geodesically convex optimization, invariant theory and polynomial identity testing. In *Proceedings of the 50th annual ACM SIGACT symposium on theory of computing*, pages 172–181, 2018.
- [Bou23] Nicolas Boumal. *An introduction to optimization on smooth manifolds*. Cambridge University Press, 2023.

- [YHAG20] Xinru Yuan, Wen Huang, P-A Absil, and Kyle A Gallivan. Computing the matrix geometric mean: Riemannian versus euclidean conditioning, implementation techniques, and a riemannian bfgs method. *Numerical Linear Algebra with Applications*, 27(5):e2321, 2020.