黎曼流形上的优化方法

张晋

2021 年 7 月 30 日

目录

1 Introduction

1.1 问题描述

对于两个 $\mathbb{R}^3$ 中的对应点集 $\mathcal{P} = \{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n\}$ 和 $\mathcal{Q} = \{\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_n\}$，我想找到一个旋转变换使得两个点集在 L1 范数的意义下配准（对于一般的点云配准来说，此处还需要平移变换，但为简化起见这里省略）。用数学方式来表示，就是想找到一个旋转矩阵 $R^\star$ 使得：

$$R^\star = \underset{R \in \mathrm{SO}(3)}{\arg\min} \sum_{i=1}^{n} \|Rp_i - q_i\|_1 \tag{1}$$

矩阵 R 如果属于旋转矩阵 SO(3)，则需要满足的条件为：

$$RR^T = I_3, \det(R) = 1$$

可将其写成矩阵的形式

$$R^\star = \underset{R \in \mathrm{SO}(3)}{\arg\min} \|RP - Q\|_1 \tag{2}$$

其中 $P = [\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_n] \in \mathbb{R}^{3\times n}, Q = [\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_n] \in \mathbb{R}^{3\times n}$。这里的 $\|\cdot\|_1$ 表示矩阵所有元素的绝对值之和。

1.2 流形优化介绍

如果没有 SO(3) 这个约束条件，那么我们可以直接计算目标函数的次梯度，然后使用梯度下降方法求解即可。

$$f(R) = \|RP - Q\|_1 \tag{3}$$

$$\nabla_R f = \mathrm{sign}(RP - Q) \cdot P^\top \tag{4}$$

但问题在于我们的 R 需要满足 SO(3) 约束，单纯使用梯度下降并不能保证满足约束条件。但幸运的是，SO(3) 是欧式空间 $\mathbb{R}^9$ 中的光滑子流形，我们可以将每一步的迭代点都限制在流形上，然后就可以在流形空间中使用那些无约束优化方法了，这就是流形优化。

直观的解释可以参考下图，假设约束流形为 $\mathcal{M}$，已知当前迭代点为 x_k，我们先计算目标函数在欧式空间的梯度 $-\nabla f(x_k)$，但新迭代点 $x_k - \nabla f(x_k)$ 不一定在流形 $\mathcal{M}$ 上，因此我们需要考虑如何构造一个合适的映射，将欧式空间的中的点映射到 $\mathcal{M}$。

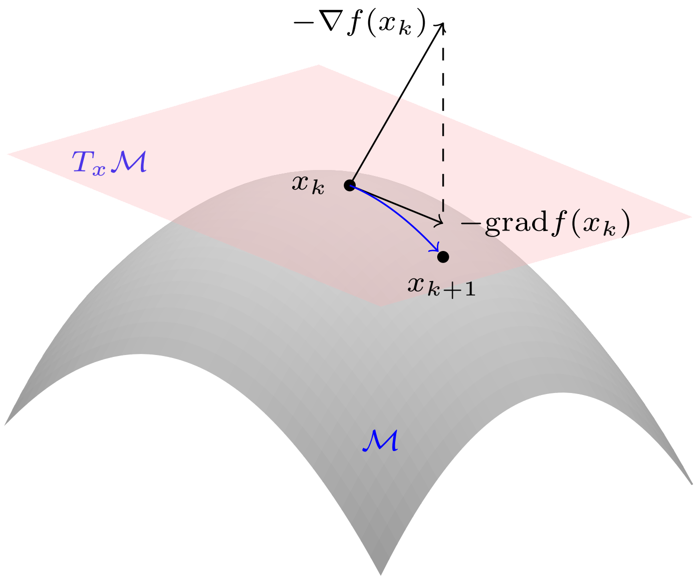

图 1: 流形优化示意图

首先考虑到切空间 $T_{x_k}\mathcal{M}$ 是流形 $\mathcal{M}$ 在点 x_k 处的线性近似，因此一个很自然的想法就是先将欧式空间的点投影到切空间 $T_{x_k}\mathcal{M}$ 上，然后再将 $T_{x_k}\mathcal{M}$ 上的点映射到 $\mathcal{M}$ ，这样就完成了一次黎曼梯度下降。常规的拟牛顿法、信赖域法等优化方法都能拓展到流形优化中。

现在我们大体思路已经明确，接下来就是具体计算每一步的变换。若只是快速上手应用，那么直接套公式即可，但如果想弄清楚公式的每一步是怎么来的，那么还需要一些黎曼几何的基础知识。

2 黎曼几何基础

2.1 切向量

$R(t)$ 是定义在 SO(3) 上的一条光滑曲线，如果 $R(t_0) = A$，那么 $H = \dot{R}(t_0)$ 就是点 A 处的一个切向量。由于 $R(t)^T R(t) = I$ ，于是有：

$$\dot{R}(t)^T R(t) + R(t)^T \dot{R}(t) = 0$$

也就是说在 SO(3) 流形中，点 A 处的切向量 H 需要满足 $H^T A + A^T H = 0$。

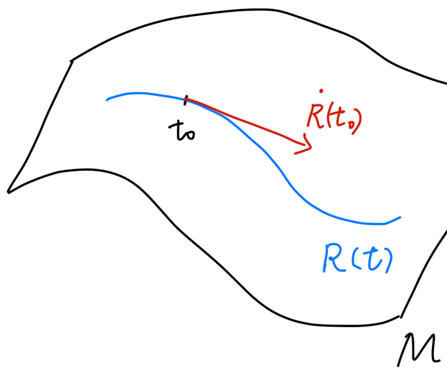

图 2: 示意图

2.2 切空间

所有曲线在点 A 处的切向量构成一个切空间 $T_A\mathrm{SO}(3)$，其定义为:

$$T_A\mathrm{SO}(3)=\{H:H^TA+A^TH=0\}$$

设 $\Omega=A^TH$，那么也可写成:

$$T_A\mathrm{SO}(3)=\{A\Omega:\Omega+\Omega^T=0\}$$

注意此处的 Ω 是一个反对称矩阵。我们知道每一个旋转矩阵都能表示为一个反对称矩阵的指数映射，而此处点 A 所对应的反对称矩阵恰恰就是 Ω，这一点我们会在后面提到。

2.3 内积

在定义了切空间后，我们需要在流形的切空间上定义内积。欧式空间中内积的定义为 $\langle a,b\rangle=a^Tb$，很自然地，我们可以将该定义沿用到子流形 SO(3) 上，切空间 $T_X\mathrm{SO}(3)$ 上的内积可定义为

$$\langle A,B\rangle_X:=\operatorname{vec}(A)^T\operatorname{vec}(B)=\operatorname{tr}\left(A^TB\right)$$

在切空间中定义了内积后，自然便有了黎曼度量，该流形也就可以称作黎曼流形了。

2.4 方向导数

对于光滑的实值函数 $f:\mathrm{SO}(3)\to\mathbb{R}$，我们可以在切空间上定义 f 沿着切向量 $H\in T_X\mathrm{SO}(3)$ 的方向导数:

$$\mathrm{D}f(X)[H]=\lim_{t\to0}\frac{f(X+tH)-f(X)}{t}$$

2.5 黎曼梯度

对于黎曼流形 $\mathcal{M}$ ，函数 f 在 x 点处的黎曼梯度 $\mathrm{grad}f(x)$ 是 $T_x\mathcal{M}$ 中的一个向量，由条件(5)唯一确定。

$$\langle \mathrm{grad}f(x), \xi \rangle_x = \mathrm{D}f(x)[\xi], \quad \forall \xi \in T_x\mathcal{M} \tag{5}$$

黎曼梯度的方向也是使方向导数最大化的方向。

由于 SO(3) 是欧式空间中的黎曼子流形，因此其黎曼梯度就是欧式空间中的梯度 $\nabla f(X)$ 在切空间 $T_X\mathrm{SO}(3)$ 上的投影：

$$\mathrm{grad}f(X) = \mathcal{P}_{T_X\mathrm{SO}(3)}(\nabla f(X))$$

投影的计算公式为：

$$\mathcal{P}_{T_X\mathrm{SO}(3)}(M) = X\mathrm{skew}\left(X^TM\right),$$

其中 skew 是反对称算子：$\mathrm{skew}(A) = \frac{A-A^T}{2}$。

最终算得：

$$\begin{aligned}\mathrm{grad}f(X) &= \frac{1}{2}\left(\nabla f(X) - X\nabla^T f(X)X\right) \\ &= \nabla f(X) - X\mathrm{sym}\left(X^T\nabla f(X)\right)\end{aligned} \tag{6}$$

其中 $\mathrm{sym}(A) = \frac{A+A^T}{2}$。

2.6 黎曼 Hessian

对于黎曼流形 $\mathcal{M}$ 上的实值函数 $f:\mathcal{M}\to\mathbb{R}$ ，给定黎曼联络 $\tilde{\nabla}$ ，可计算其黎曼 Hessian：

$$\mathrm{Hess}f(x)[\xi] := \tilde{\nabla}_\xi \mathrm{grad}f(x), \quad \xi \in T_x\mathcal{M}$$

同样的，要计算 SO(3) 的黎曼 Hessian，只需将其黎曼梯度在欧式空间中的方向导数投影到切空间上即可：

$$\begin{aligned}\mathrm{Hess}f(X)[U] &= \mathcal{P}_{T_X\mathrm{SO}(3)}(D\mathrm{grad}f(X)[U]) \\ &= \mathcal{P}_{T_X\mathrm{SO}(3)}\left(\nabla^2 f(X)[U] - U\mathrm{sym}\left(X^T\nabla f(X)\right) - XS\right) \\ &= \mathcal{P}_{T_X\mathrm{SO}(3)}\left(\nabla^2 f(X)[U] - U\mathrm{sym}\left(X^T\nabla f(X)\right)\right)\end{aligned}$$

其中 $S = \mathrm{sym}\left(U^T\nabla f(X) + X^T\nabla^2 f(X)[U]\right)$ ，$U \in T_X\mathrm{SO}(3)$ ，可算出 XS 在切空间上的投影为 0。

2.7 测地线

测地线是黎曼流形上连接两点之间的最短曲线，并且测地线的沿曲线加速度 $\gamma''(t)$ 恒为 0。沿曲线加速度可以通过曲线在欧式空间中加速度的投影得到：

$$\gamma''(t) = \mathcal{P}_{T_{\gamma(t)}\mathcal{M}}(\ddot{\gamma}(t))$$

也就是说，测地线在欧式空间中的加速度方向与切空间垂直。

2.8 指数映射

我们现在要计算切空间 $T_x\mathcal{M}$ 到流形 $\mathcal{M}$ 上的映射 f ，一个很自然的想法就是将切向量 v 投影到流形上，那么哪些条件需要被满足呢？首先，切空间的原点必须映射到切点，即 $f(0) = x$；然后直线 vt 在流形上的投影必然是一条测地线 $r(t)$；如果切向量 v 的长度变成了原来的 a 倍，那么测地线的长度也应该要增大到原来的 a 倍。我们将上面这些要求整理之后，得到如下条件：

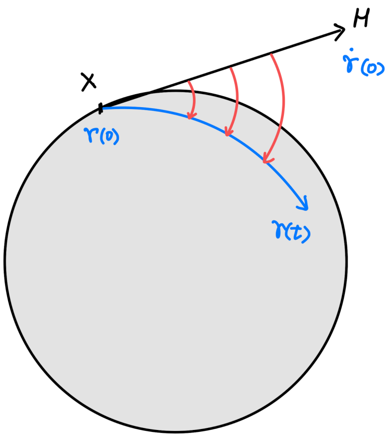

图 3: 示意图

对于黎曼流形 $\mathcal{M}$ 上的一个点 x ，给定 x 点处的一个切向量 v ，寻找满足以下 3 个条件的测地线 $\gamma(t)$：

1. $\gamma(0) = x$
2. $\dot{\gamma}(0) = v$
3. $\gamma(at; x, v) = \gamma(t; x, av)$

那么可以定义从切空间 $T_x\mathcal{M}$ 到流形 $\mathcal{M}$ 上的映射：

$$\mathrm{Exp}\left(x,v\right)=\mathrm{Exp}_x\left(v\right)=\gamma\left(1;x,v\right) \tag{7}$$

该映射就称为在 x 点处的指数映射。

对于 SO(3) 流形，以 X 为起点，沿切向量 H 方向延伸的测地线方程如下，具体推导过程见附录。

$$\gamma\left(t\right)=X\exp\left(tX^TH\right) \tag{8}$$

由此可知 SO(3) 在 X 处的指数映射为：

$$\mathrm{Exp}_X\left(H\right)=X\exp\left(X^TH\right) \tag{9}$$

2.9 对数映射

对数映射从流形映射到切空间，是指数映射的逆，SO(3) 在 X 处的对数映射为：

$$\mathrm{Log}_X\left(Y\right)=X\log\left(X^TY\right) \tag{10}$$

2.10 测地距离

对于点 Y，若 $\mathrm{Log}_X\left(Y\right)=V$，那么有 $\mathrm{Exp}_X\left(V\right)=Y$，我们可以写出 X 到 Y 的测地线方程：

$$\gamma(t)=X\exp\left(tX^TV\right) \tag{11}$$

其中 $\gamma(0)=X,\gamma(1)=Y$，那么 X 到 Y 的测地距离为：

$$\begin{aligned}\mathrm{dist}(X,Y)&=\int_0^1\|\gamma'(t)\|_{\gamma(t)}\,\mathrm{d}t\\&=\int_0^1\|\gamma(t)X^TV\|_{\gamma(t)}\,\mathrm{d}t\\&=\int_0^1\|V\|_{\gamma(t)}\,\mathrm{d}t\\&=\|V\|\end{aligned} \tag{12}$$

由此可见 V 的模长 $\|V\|$ 就是 X 到 Y 的测地距离。

我们闲得无聊可以继续往下算，先设 $\exp\left(\theta[n]_\times\right)=X^TY$，那么有：

$$\begin{aligned}\mathrm{dist}(X,Y)^2&=\langle X\theta[n]_\times,X\theta[n]_\times\rangle\\&=\theta^2\mathrm{tr}\left([n]_\times^T\cdot[n]_\times\right)\\&=2\theta^2\end{aligned} \tag{13}$$

也就是说 X 到 Y 的测地距离为 $\sqrt{2}\theta$，这里的 θ 是矩阵 X^TY 的旋转角。

3 流形优化算法

3.1 黎曼最速下降法

黎曼最速下降法以 x_k 处的负黎曼梯度方向为搜索方向 $\eta_k = -\mathrm{grad} f\left(x_k\right)$，然后通过线搜索寻找合适的步长 t_k，最后通过指数映射将切向量 $t_k\eta_k$ 映射到流形 $\mathcal{M}$ 上，作为下一个迭代点 $x_{k+1} = \mathrm{Exp}_{x_k}(t_k\eta_k)$。

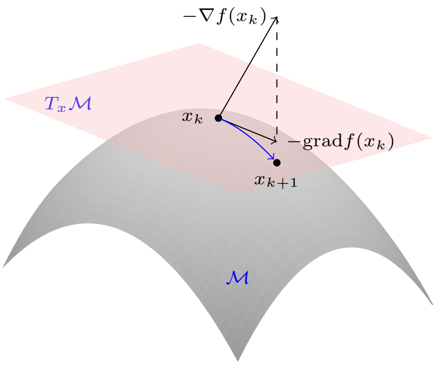

图 4: 黎曼最速下降法

Algorithm 1 黎曼最速下降法

Construct an initial $x_0 \in \mathcal{M}$
for $k = 0, 1, 2, \ldots$ **do**
  $\eta_k = -\mathrm{grad} f\left(x_k\right)$ ▷ 计算搜索方向
  Choose $t_k > 0$ ▷ 计算 Armijo 步长
  $x_{k+1} = \mathrm{Exp}_{x_k}(t_k\eta_k)$ ▷ 映射回流形上
end for

3.2 黎曼牛顿法

黎曼流形上的牛顿迭代步如下：

$$x_{k+1} = x_k - \left(\mathrm{Hess}\, f\left(x_k\right)\right)^{-1}\left[\mathrm{grad}\, f\left(x_k\right)\right] \tag{14}$$

3.3 原问题求解

由于原问题中的 Hess 阵恒为 0，因此不能直接使用牛顿法，这里我们采用拟牛顿法进行求解。此外为了加快算法的收敛性，我们用 Huber 函数近似代替绝对值函数。该程序可通过调用 Manopt 优化库 实现。

4 附录

4.1 测地线的计算

计算 SO(3) 流形上以 X 为起点，沿切向量 H 方向延伸的测地线方程。

$$\gamma^T(t)\cdot\gamma(t)=I \tag{15}$$

$$\dot{\gamma}^T(t)\cdot\gamma(t)+\gamma^T(t)\cdot\dot{\gamma}(t)=0 \tag{16}$$

设 $\Omega(t)=\gamma^T(t)\cdot\dot{\gamma}(t)$ ，显然 $\Omega(t)$ 是反对称矩阵，也有：

$$\dot{\gamma}(t)=\gamma(t)\cdot\Omega(t) \tag{17}$$

$\dot{\gamma}(t)$ 显然是属于切平面 $T_X\mathrm{SO}(3)$ 的，因此沿曲线导数就是它本身：

$$\gamma'(t)=\mathcal{P}_{T_{\gamma(t)}\mathcal{M}}\left(\dot{\gamma}\left(t\right)\right)=\dot{\gamma}(t) \tag{18}$$

于是可以计算：

$$\begin{aligned}\gamma''(t)&=\mathcal{P}_{T_{\gamma(t)}\mathcal{M}}\left(\frac{\mathrm{d}}{\mathrm{d}t}\gamma'(t)\right)\\&=\mathcal{P}_{T_{\gamma(t)}\mathcal{M}}\left(\ddot{\gamma}\left(t\right)\right)\\&=\mathcal{P}_{T_{\gamma(t)}\mathcal{M}}\left(\dot{\gamma}(t)\cdot\Omega(t)+\gamma(t)\cdot\dot{\Omega}(t)\right)\end{aligned} \tag{19}$$

上文提到过投影的计算公式为：

$$\mathcal{P}_{T_X\mathrm{SO}(3)}\left(M\right)=X\mathrm{skew}\left(X^TM\right)$$

可计算得：

$$\begin{aligned}&\gamma^T(t)\left(\dot{\gamma}(t)\cdot\Omega(t)+\gamma(t)\cdot\dot{\Omega}(t)\right)\\&=\gamma^T(t)\cdot\dot{\gamma}(t)\cdot\Omega(t)+\gamma^T(t)\cdot\gamma(t)\cdot\dot{\Omega}(t)\\&=\Omega^2(t)+\dot{\Omega}(t)\end{aligned} \tag{20}$$

由于 $\Omega^2(t)$ 为对称阵，$\dot{\Omega}(t)$ 为反对称阵，因此有：

$$\gamma''(t)=\gamma(t)\mathrm{skew}\left(\Omega^2(t)+\dot{\Omega}(t)\right)=\gamma(t)\cdot\dot{\Omega}(t) \tag{21}$$

$$\dot{\Omega}(t)=\gamma^T(t)\gamma''(t) \tag{22}$$

由于 $\gamma(t)$ 为测地线，有 $\gamma''(t)=0$ ，于是可推导出：

$$\Omega(t)=\gamma^T(0)\cdot\dot{\gamma}(0)=X^TH \tag{23}$$

解微分方程 $\gamma(t)\cdot\Omega=\dot{\gamma}(t)$ 得：

$$\gamma\left(t\right)=X\exp\left(tX^TH\right) \tag{24}$$

5 相关书籍推荐

[Hu et al., 2019] 是袁亚湘团队对流形优化的一个简要介绍，可以看一看作为入门。

[Boumal, 2020] 最近新出的一本介绍流形优化的书，排版精美内容详实，如果想深入了解的话强烈推荐看这本书。

[Absil et al., 2008] 为数不多介绍矩阵流形的书，也可以看一看。

[Boumal and Absil, 2011] 主要是 SO(3) 上的流形优化。

[Solà et al., 2020] 这个从机器人应用角度来介绍 SO(3) 上的李群理论。

[Boumal et al., 2013] 是一个 MATLAB 的流形优化库，可以快速上手。

参考文献

[Absil et al., 2008] Absil, P.-A., Mahony, R., and Sepulchre, R. (2008). *Optimization algorithms on matrix manifolds*. Princeton University Press, Princeton, N.J. ; Woodstock. OCLC: ocn174129993.

[Boumal, 2020] Boumal, N. (2020). An introduction to optimization on smooth manifolds. *Available online, Aug*, page 310.

[Boumal and Absil, 2011] Boumal, N. and Absil, P. A. (2011). A discrete regression method on manifolds and its application to data on SO(n). *IFAC Proceedings Volumes*, 44(1):2284–2289.

[Boumal et al., 2013] Boumal, N., Mishra, B., Absil, P.-A., and Sepulchre, R. (2013). Manopt, a Matlab toolbox for optimization on manifolds. *arXiv:1308.5200 [cs, math, stat]*. arXiv: 1308.5200.

[Hu et al., 2019] Hu, J., Liu, X., Wen, Z., and Yuan, Y. (2019). A Brief Introduction to Manifold Optimization. *arXiv:1906.05450 [math]*. arXiv: 1906.05450.

[Solà et al., 2020] Solà, J., Deray, J., and Atchuthan, D. (2020). A micro Lie theory for state estimation in robotics. *arXiv:1812.01537 [cs]*. arXiv: 1812.01537.