arXiv:2602.05174v1 [stat.ML] 5 Feb 2026

黎曼流匹配的全变差收敛速率

Yunrui Guan[1],Krishnakumar Balasubramanian[2],和 Shiqian Ma[1]

[1]计算应用数学与运筹学系，莱斯大学。[2]统计系，加利福尼亚大学戴维斯分校。
[1]{yg83,sqma}@rice.edu[2]{kbala}@ucdavis.edu

摘要

黎曼流匹配（RFM）通过学习时变切向量场，将基于流的生成建模扩展到支撑于流形上的数据；该向量场所对应的流常微分方程将简单的基分布输运至数据分布。我们针对使用学习向量场并在流形上采用欧拉离散化的黎曼流匹配采样器，建立了非渐近的全变差（TV）收敛性分析。我们的关键技术工具是一个刻画两个流形常微分方程流之间全变差演化的微分不等式，它通过向量场失配的散度和参考流的得分函数来表示全变差的时间导数；控制这些项需要建立能够明确考虑平行移动和曲率的新界。在总体流匹配向量场满足光滑性假设，且学习场具有均匀（紧流形）或均方（Hadamard 流形）逼近保证的条件下，我们得到如下形式的显式界 TV $\le C_{\mathrm{Lip}}\, h + C_\varepsilon\, \varepsilon$（在紧流形上还包含一个高阶 ε^2 项），并清晰地区分数值离散化误差与学习误差。这里，h 是步长，ε 是目标精度。具体应用可在超球面上得到显式多项式迭代复杂度 S^d；在满足温和矩条件的对称正定矩阵（n）流形上也可得到相应结果。

目录

1 引言 **2 2 基础知识与问题表述 4**

2.1 黎曼几何基础.2.2 黎曼流匹配. 4
5

3 全变差收敛速率：紧流形 7

4 全变差收敛速率：Hadamard 流形 9

5 证明概要与中间结果 10
5.1 连续时间插值.115.2 离散化分析.125.2.1 得分正则性与向量场正则性.125.2.2 散度项.13

A 黎曼几何补充背景 **17**
A.1 使用的距离.17A.2 比较定理.17A.3 黎曼子流形.18

B 主定理证明 定理 18
B.1 引理证明 5.1 .19B.2 速度向量项.20B.3 散度项.23B.4 黎曼流形上的结果.26B.5 主定理.
30B.6 示例：超球面.32B.7 示例：对称正定矩阵流形.35

C 主要定理证明的辅助结果 36
C.1 雅可比方程.36C.2 散度项.39

D 超球面 正则性结果 4 **2**
D.1 辅助引理.45D.2 流匹配向量场的正则性.53D.3 散度的正则性.56D.4 v(t, x) 和 log pt 的正则性.62D.5 得分正则性的有限性：命题5.3 的证明.63

E 对称正定矩阵流形正则性结果 65
E.1 命题证明 5.4 .68E.2 Hadamard 流形和 SPD 流形正则性的辅助结果.70E.2.1 期望控制.
70E.2.2 正则性控制.73E.3 导数的辅助界.79E.4 指数映射三阶导数的界.87E.5 涉及黎曼对数映射的界.96E.6 辅助结果.99

1 引言

基于流的生成建模提供了一种概念上清晰的采样范式，即将复杂的目标分布表示为从简单参考概率分布出发的连续时间输运的终点。在流匹配中，人们固定一族插值测度（π_t）$_{t\in[0,1]}$ 介于易于处理的基分布 π_0 与数据分布 π_1之间，并学习一个时变向量场，其流将 π_0 推动至 π_1。当数据位于黎曼流形（M， g）上时，粒子按照内在常微分方程 $\dot{X}_t = v_t$（X_t）演化，相应的密度满足流形连续性方程$\partial_t p_t$ +散度$_g$（$p_t v_t$） $= 0$，从而得到一个在构造上尊重几何的采样器。关键在于，由于黎曼流匹配（RFM）是纯确定性的，并且仅依赖于对内在常微分方程进行积分，它避免了黎曼扩散模型所要求的对流形值布朗运动进行离散化（De Bortoli 等人，2022年），而在这些模型中，理论分析与实际模拟都要精细得多。

尽管它们的实际应用日益广泛（Yue 等人，2025年；Sriram 等人，2024年；Mathieu 和 Nickel，2020；Miller等人，2024年；Collas等人，2025年；Luo等人，2025年），一个核心理论问题仍未得到探讨：带有学习到的速度场的离散黎曼流匹配采样器，在分布度量下以多快的速度收敛到目标分布？我们的主要贡献是为紧流形与Hadamard流形上的黎曼流匹配采样器提供首个非渐近全变差（TV）界。我们建立了驱动向量场受扰动时黎曼连续性方程解的内在稳定性估计，从而对$T = 1 - \delta$、（$\delta > 0$）情形下的TV($\hat{}\pi_T, \pi_T$)给出显式控制，其依赖于：(i) 学习场$\hat{}v_t$在几何适配范数下对理想场 v_t的逼近质量；(ii) 动力学的正则性（例如对协变导数与散度的界）；以及 (iii) 支配体积畸变的（M, g）的几何特征。随后，我们将这些稳定性估计与对离散流的细致分析相结合，为实用采样器获得端到端保证，并分离出统计贡献与计算贡献。所得理论阐明了学习到的速度场的哪些性质对于流形上的强采样精度真正是必需的，以及曲率与各向异性如何进入常数与速率，从而为理解黎曼流匹配算法提供了原则性基础。

相关工作。De Bortoli 等人将黎曼流形上的生成建模引入De Bortoli 等人(2022年)，他们将基于得分的生成建模（Song等人，2021）从欧几里得空间推广到黎曼流形，并证明了采样误差的误差界。De Bortoli 等人De Bortoli 等人(2022年的误差界并非多项式级，这源于分析流形布朗运动模拟中的离散化误差时遇到的技术困难。其他工作，包括Huang等人(2022年）和Lou等人(2023年），也研究了流形上的生成建模，但没有给出理论保证。

这一困难与通过扩散过程在黎曼流形上进行采样有关。早期工作分析了朗之万扩散过程中的 KL 散度，例如超球面（Li和Erdogdu，2023年）和海森矩阵流形（Gatmiry和Vempala，2022年）。随后，Cheng 等人(2022年采用结合耦合的几何方法，得到了 W_1 界；关于李群，另见Kong和Tao(2024年)，关于高精度近端采样，另见Guan 等人(2025年)。这些结果表明，作为黎曼扩散模型基本组成部分的黎曼布朗运动，无论在理论上还是计算上都颇具挑战：离散化分析大多依赖几何方法，通常给出 Wasserstein误差界（例如，Cheng 等人(2022年）），而由于转移密度除特殊情形外没有闭式表达，通常无法进行精确模拟。这也使黎曼扩散模型中的去噪分数匹配和样本生成更加复杂。

近年来，Xu 等人(2026年，引理 19）确立了全变差距离下的多项式离散化误差界。在此基础上，Xu 等人(2026年) 进一步证明了黎曼分数生成模型具有多项式迭代复杂度。然而，现有界限仅具定性，未能揭示其对问题参数的精确依赖关系。

另一条研究路线是黎曼流匹配方法，该方法最初由Chen 和 Lipman(2024年) 提出，随后由 Cheng 等人(2025年) 和Wu 等人(2025年) 进一步研究。我们指出，由于其确定性表述，RFM 无需获取热核或模拟流形布朗运动，从而避免了黎曼扩散模型的主要瓶颈。

从欧几里得流匹配的情况来看，包括 Li 等人 (2025年) 和Huang等人(2025年) 在内的研究，为概率流常微分方程建立了全变差距离下的离散化误差界，其对利普希茨常数呈多项式依赖。对于通过常微分方程离散化实现确定性采样的流匹配，Benton 等人 (2024年)；Bansal 等人 (2024年)；Zhou 和 Liu (2025年)；Guan 等人(2026年) 建立了 Wasserstein距离下的误差界，而 Su 等人 (2025年) 研究了 KL散度。然而，如果得分函数或流匹配向量场不具备收缩/耗散型结构，这些误差界通常无法给出关于相应 Lipschitz常数的多项式速率。为了

获得全多项式误差界，现有研究（Li 等人，2024a，2025年）在全变差距离下分析流匹配采样器。Liu 等人(2025年) 利用 Li 等人 (2025年) 的框架，为随机插值器（可理解为流匹配的一种平滑变体）的确定性采样给出了多项式误差界。最近，Roy 等人 (2026年) 研究了欧几里得流匹配对低维结构的自适应性。据我们所知，目前尚无针对一般黎曼流形上流匹配的离散化误差分析。

2 基础知识与问题表述

2.1 黎曼几何基础

设 (M, g) 为一个 d 维黎曼流形，该流形不一定被实现为欧几里得空间中的子流形，其中 g 表示黎曼度量。除非另有说明，$\|\cdot\|$ 表示由 g 诱导的范数。我们用 T_xM 表示位置 $x \in M$ 处的切空间，并注意到 $\|\cdot\|$ 在每个切空间 T_xM 上都有明确定义。我们用 Exp 表示指数映射，并定义指数映射的逆映射，即对数映射：给定 $X_0, X_1 \in M$，$\mathrm{Log}_{X_0}(X_1) \in T_{X_0}M$ 满足 $\mathrm{Exp}_{X_0}(\mathrm{Log}_{X_0}(X_1)) = X_1$。除非另有说明，我们用 P_x^y 表示沿最短测地线从 x 到 y 的平行移动。在此基础上，我们可以如下定义测地线插值。

$$X_t := \mathrm{Exp}_{X_0}(t\,\mathrm{Log}_{X_0}(X_1)) \tag{1}$$

注意，$X_t' = P_{X_0}^{X_t}\mathrm{Log}_{X_0}(X_1) = \frac{1}{1-t}\mathrm{Log}_{X_t}(X_1)$。

我们用 grad 表示黎曼梯度，用 div 表示黎曼散度，并用 ∇ 表示 (M, g) 上的 Levi–Civita 联络。对于光滑函数 f，梯度 grad f 是满足以下条件的唯一向量场：对所有 $v \in T_xM$，均有 $\langle \mathrm{grad}\ f, v\rangle = df(v)$。对于光滑向量场 u，散度定义为 $\mathrm{div}\ u := \mathrm{tr}(\nabla u)$；等价地，对于任意局部 g-正交规范标架 $\{e_i\}_{i=1}^d$，有 $\mathrm{div}\ u(x) = \sum_{i=1}^d \langle \nabla_{e_i}u, e_i\rangle_x$。对于向量场 u 和切向量 $v \in T_xM$，协变导数 $\nabla_v u(x)$ 定义如下：任取满足 $V(x) = v$ 的光滑局部向量场 V，并令 $\nabla_v u(x) := (\nabla_V u)(x)$；这是良定义的（即与延拓 V 的选取无关），因为 ∇ 在第一个自变量上是 $C^\infty(M)$-线性的。映射 $\nabla u(x)$ 可视为一个 (1, 1) 张量（自同态）$\nabla u(x) : T_xM \to T_xM$，其定义为 $(\nabla u(x))(v) := \nabla_v u(x)$。因此，我们定义由 g 诱导的算子范数为

$$\|\nabla u(x)\|_{\mathrm{op}} := \sup_{v\neq 0}\frac{\|\nabla_v u(x)\|}{\|v\|} = \sup_{\|v\|=1}\|\nabla_v u(x)\|.$$

现在我们来直观介绍 Levi–Civita 联络和协变导数。在流形上，不同点处的切空间 T_xM 和 T_yM 是不同的向量空间，因此无法直接相减位于 x 和 y 处的向量。Levi–Civita 联络 ∇ 提供了一种规范方式，通过比较附近的切向量来微分向量场，同时与度量（$\nabla g = 0$）相容且无挠率。沿着曲线 γ，诱导出的协变导数$D_t := \nabla_{\gamma'(t)}$ 起到方向导数的作用：它在考虑切空间变化之后，衡量向量场沿着 γ 的内禀变化率；而正是这种导数概念，使得沿着 γ 的“速度恒定”与测地线方程相一致。

设 $x \in M$。x 处的单射半径记为 $\mathrm{inj}(x)$，定义为

$$\mathrm{inj}(x) := \sup\Big\{a > 0 : \mathrm{Exp}_x\big|_{B_a(0)} : B_a(0) \subset T_xM \to M \text{ is a diffeomorphism onto its image}\Big\},$$

其中 $B_a(0)$ 是 $(T_xM,\ g_x)$ 中的开度量球。 M 的单射半径为 $\mathrm{inj}(M) := \inf_{x\in M} \mathrm{inj}(x)$。

令 R 表示与 ∇ 相关的（黎曼）曲率张量，其对光滑向量场 $X,\ Y,\ Z$定义为 $R(X,\ Y)Z$: $= \nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z - \nabla_{[X,Y]}Z$。对于线性无关的 $u,\ v \in T_xM$，平面 $\sigma = \mathrm{span}\{u, v\}$ 的截面曲率为

$$K(\sigma) := \frac{\langle R(u,v)v,\ u\rangle_x}{\|u\|^2\|v\|^2 - \langle u, v\rangle_x^2},$$

特别地，如果 $u \perp v$ 和 $\|u\| = \|v\| = 1$，那么 $K(u,\ v) = \langle R(u,\ v)v,\ u\rangle_x$.

设 $\gamma\colon [a,b] \to M$ 是一条测地线，并记 $D_t := \nabla_{\gamma'(t)}$ 为沿 γ 的协变导数。沿 γ 的向量场 J 若满足雅可比方程 $D_t^2 J + R(J,\ \gamma')\gamma' = 0$，则称为雅可比场。令 $p = \gamma(a)$ 和 $q = \gamma(b)$。如果存在一个非零的雅可比场 J，沿 γ 且满足 $J(a) = J(b) = 0$，则称 p 和 q 沿共轭于 γ。

对于单位速度测地线 $\gamma\colon [0,\infty) \to M$，若 $\gamma(0) = p$，定义其截断时刻 $t_{\mathrm{cut}}(\gamma) := \sup\{t > 0 :$ $\gamma|_{[0,t]}$ 其端点之间的距离达到最小值$\}$. 若 $t_{\mathrm{cut}}(\gamma) < \infty$，则沿 p 的割点 γ 为$\gamma(t_{\mathrm{cut}}(\gamma))$。 p的割迹，记为 $\mathrm{Cut}(p)$，是从 p出发的所有单位速度测地线上 p 的割点构成的集合。一个经典事实是，沿任意一条从 p出发的测地线，截断时刻不晚于首次共轭时刻（若存在共轭点）；此外，当同一端点存在多条最短测地线时，割点可能严格早于首次共轭点出现。

现在我们对上述概念作一些直观说明。曲率张量刻画了协变导数不满足交换律的程度；它编码了流形的内在弯曲。截面曲率将这些信息简化为二维方向 σ：正截面曲率往往使 σ 中的邻近测地线彼此靠拢，而负曲率则往往使它们相互远离。雅可比场描述邻近测地线之间的无穷小分离（它们由测地线的变分产生），因此雅可比场的零点对应于测地线局部唯一性或最短性的丧失。共轭点描述一种微分性质，对应于 d指数映射的退化，因此自然会出现在我们后续分析所用的比较理论中（见第 A.2节）。另一方面，割点描述一种几何性质，说明测地线何时不再是最短测地线，以及对数映射$_x$ 何时不再是单值的。

2.2 黎曼流匹配

设 π_0 和 π_1 是支撑在黎曼流形 M上的两个概率分布。黎曼流匹配（RFM）框架旨在学习一个时变向量场 $v(t,\ x)$，其诱导的流将样本从 π_0 输送到 π_1。具体而言，RFM 寻求 v，使得若 $X_0 \sim \pi_0$，则常微分方程的解 $\{X_t\}_{t\in[0,1]}$

$$dX_t = v(t, X_t)\,dt, \qquad X_0 \sim \pi_0, \tag{2}$$

满足 $X_1 \sim \pi_1$。只要 v 关于时间连续且关于空间局部利普希茨连续，常微分方程就存在极大解；例如参见（Lang，2012年，第四章）。此外，如果流形是完备的，且 $v(t,\ x)$ 满足线性增长界，则解在全局时间范围内存在。这里还要指出，Wan等人 (2025年) 针对欧几里得空间的情形给出了更精细且适用范围更广的条件。

一个关键的几何约束是，对于每个 (t,x)，速度必须位于当前点处的切空间：

$$v(t,x) \in T_xM, \qquad t \in [0,1].$$

也就是说，$v(t,x)$ 表示沿流形 在 x处一个有效的瞬时运动方向。这正是黎曼设定与欧氏设定存在根本差异之处。当 $M=\mathbb{R}^d$时，所有切空间在规范意义下都相同，即对于所有 x， $T_x\mathbb{R}^d \cong \mathbb{R}^d$，因此我们可以将向量场简单地视为一个全局映射 $v\colon [0,1]\times\mathbb{R}^d \to \mathbb{R}^d$，其陪域是一个固定的向量空间。

然而，在一般流形 M 上，切空间 $\{T_xM\}_{x\in M}$构成了一族附属于不同点的不同向量空间。虽然每个 T_xM 的维数都相同，但当 $x\neq y$ 时， T_xM 和 T_yM 之间并不存在规范的对应关系。因此，$v(t,\ \cdot)$ 的"输出空间"取决于 x： v 是切丛的一个截面，而非映射到某个固定向量空间$\mathbb{R}^d$的映射。这种位置依赖性是欧几里得流匹配与 RFM 之间技术差异的主要来源，后续分析中也会涉及这一点。

在流匹配中，通常使用神经网络对时变向量场进行参数化，并通过最小化条件流匹配 目标函数来学习该向量场。在黎曼设定下，若选择端点之间的耦合为独立耦合，即独立采样$X_0\sim\pi_0$ 和 $X_1\sim\pi_1$，则训练损失为

$$\min_u\ \mathbb{E}_{t,\,X_0\sim\pi_0,\,X_1\sim\pi_1}\left[\|u(X_t,t)-P_{X_0}^{X_t}\operatorname{Log}_{X_0}(X_1)\|^2\right], \tag{3}$$

其中 $\operatorname{Log}_x(\cdot)$ 表示黎曼对数映射，而 P_a^b表示沿插值曲线从 a 到 b的平行移动。在欧几里得空间中，标准选择是直线插值 $X_t:=tX_1+(1-t)X_0$，位于 X_0 和 X_1之间；在黎曼流形上，这自然推广为式 X_t （1）中的测地线插值。我们将总体极小化子记为 $v(t,\ x)$，将学习得到的（神经网络）近似记为 $\hat{}v(t,\ x)$。

值得注意的是，总体极小化子可表示为一个条件期望的闭式表达式（例如，参见 Guan 等人 (2026年）。具体而言，对于任意固定的 $t<1$ 和 $x\in M$，

$$\begin{aligned} v(t,x) &= \mathbb{E}\left[P_{X_0}^{x}\ \operatorname{Log}_{X_0}(X_1)\,\middle|\,X_t=x\right] = \frac{1}{1-t}\,\mathbb{E}[\operatorname{Log}_x(X_1)\,|\,X_t=x] \\ &= \frac{1}{1-t}\int_M \operatorname{Log}_x(x_1)\,p_t(x_1\mid x)\,dV_g(x_1), \end{aligned} \tag{4}$$

其中， $p_t(x_1\mid x)$ 表示在上述独立耦合和所选插值下，给定 $X_t=x$ 时 X_1 的条件密度，而 dV_g 是黎曼体积测度。

Algorithm 1 RFM with Early Stoppping

1: Denote the learned vector field by $\hat{v}(t,x)$.
2: Choose step size schedule $\{h_i\}$ for Euler discretization and early stopping time $T\in[0,1)$ such that $\sum_{i=0}^{N-1}h_i=T$, where N is the total number of steps.
3: Perform Euler discretization for N steps:
4: Generate $x_0\sim\pi_0$, set $t_0=0$.
5: **for** $k=0$ to $N-1$ **do**
6: $\quad x_{t_{k+1}}=\operatorname{Exp}_{x_{t_k}}(h_k\hat{v}(t_k,x_{t_k}))$.
7: $\quad t_{k+1}\leftarrow t_k+h_k$.
8: **end for**
9: Output a sample $x_T:=x_{t_N}\sim\hat{\pi}_T$.

为生成样本，我们通过数值模拟算法 1 中学习得到的流匹配常微分方程。尽管理想的（总体）流由真实的极小化子 v驱动，会在时刻 $t=1$将 π_0 传输至 π_1，但在实践中，我们只能使用学习得到的近似值 $\hat{}v$ 和时间离散积分器。一个关键难点是条件流匹配向量场在终止时刻附近的行为。由（4）可知，总体极小化子中显式包含一个 $1/(1-t)$ 因子。沿着用于定义 精确测地桥的 X_t，项 $\mathbb{E}$[对数映射$x(X_1)\ \mid X_t=x$] 通常按如下方式缩放

$(1-t)$，使得 $v(t,\ x)$ 保持有限；然而，这种抵消很脆弱。学习场 $\hat{}v$ 与总体场 v之间的任何不匹配，或模拟轨迹偏离

训练插值，都可能破坏这种抵消，并使有效动力学随着 $t \to 1$ 变得越来越刚性（速度较大和/或对 x 的敏感度较高）。

这种刚性直接促使我们采用提前停止。在算法 1 中，我们用 $\{h_i\}$ 表示步长调度。在恒定步长的情况下，我们只需令 $h_i \equiv h, \forall i$。给定预先设计的步长调度和早停时间 T，该算法生成 $\{x_{t_k}\}$，其中 $t_k := \sum_{i=0}^{k-1} h_i$，因此根据定义有 $t_N = T$。欧拉离散化会累积数值离散化误差，以及学习 $\hat{v}$ 所带来的统计/近似误差。在 $t = 1$附近，因子 $\frac{1}{1-t}$会放大对较小条件均值 $\mathbb{E}[\mathrm{Log}_x(X_1) \mid X_t = x]$的估计误差：该条件期望中的微小绝对误差可能转化为大得多的速度误差，进而在最后几个积分步中造成 $O(1)$ 级扰动，并使终端分布偏离 π_1。因此，遵循欧几里得空间中的现有概率流分析（例如，Li 等人（2025年，2024a）），我们建立了在终止时刻 $T < 1$收敛至近似分布的速率；此时积分会在进入条件最差的区域之前停止。

3 全变差收敛速率：紧流形

我们现在为紧流形建立全变差误差界，涵盖 SO(3)（Yue 等人，2025年）、环面和超球面（Mathieu 和 Nickel，2020；Sriram 等人，2024年）等应用。随后，我们针对超球面这一情形进行专门分析，以获得显式的迭代复杂度结果。首先给出关于流形曲率的以下假设。

假设 1 (关于黎曼流形的假设)。 M 是一个无边界的完备流形。存在 $L_R \geq 0$，使得 $\|R(u, v)v\| \leq L_R \|u\| \|v\|^2, \forall u, v, w \in T_x M$，且$M$ 的所有截面曲率均有下界 $K_{\min}$ 和上界 $K_{\max}$。

这一假设确保了 M几何结构具有标准的整体正则性：完备性保证测地线和指数映射在任意时刻都有定义。界 $\|R(u,v)v\| \leq L_R \|u\| \|v\|^2$ 控制曲率算子的范数，从而确保沿测地线的曲率诱导偏差均匀有界。最后，双侧截面曲率界 $K_{\min} \leq K \leq K_{\max}$ 可防止流形的正曲率过大（测地线过度聚焦）或负曲率过大（测地线过快发散），这对稳定性和比较估计至关重要。我们对流匹配向量场 v 作出以下假设。

假设 2. （流匹配向量场的正则性）真实向量场 $v(t, x)$ 在 t, x 中至少C^2，并对 $\forall x \in M, t \in [0, 1)$满足以下条件：

(1) $\|\nabla v(t, x)\|_{\mathrm{op}} \leq L_{v,t}^{x}$。*(2)* $\|v(t, x)\| \leq L_v$ 且 $\left\| \frac{d}{dt} v(t, x) \right\| \leq L_{v,t}^{x}$。*(3)* $\|$ 梯度 $_x$ 散度 $v(t, x)\| \leq L_{\mathrm{div},t}^{x}$ 且 $\left| \frac{d}{dt}$ 散度 $v(t, x) \right| \leq L_{\mathrm{div},t}^{x}$。 —

上述条件要求总体向量场 $v(t, x)$ 平滑且均匀地表现良好，从而使其诱导的流随时间和初始条件稳定变化。对$\|\nabla v(t, x)\|_{\mathrm{op}}$、$\|v(t, x)\|$和 $\left\| \frac{d}{dt} v(t, x) \right\|$ 的界限确保动力学在空间上满足利普希茨条件，且其幅度和时间变化受到控制，从而防止轨迹分离过快，或随着 t 演化而出现不稳定性。在具有标准度量的欧几里得空间 $M = \mathbb{R}^d$ 中，$\nabla v(t, x)$化为雅可比矩阵 $D_x v(t, x)$，因此 $\|\nabla v(t, x)\|_{\mathrm{op}}$ 恰好就是通常的谱范数／算子范数 $\|D_x v(t, x)\|_{2 \to 2}$（即 $v(t, \cdot)$ 在 x处的利普希茨常数）。对 $\|$ 梯度$_x$ 散度 $v(t, x)\|$ 和 $\left|\frac{d}{dt}$ 散度 $v(t, x)\right|$的界限 —

控制流所诱导的局部体积变化如何随空间和时间而变化，这对于分析密度如何根据连续性方程演化至关重要。

假设 3。*(估计误差)* 存在某个小的 $\varepsilon > 0$，使得以下条件成立：$\forall x \in M, t \in [0,1)$：*(1)* $\|\hat{v}(t,x) - v(t,x)\| \le \varepsilon$，以及 *(2)* $\|\nabla \hat{v}(t,x) - \nabla v(t,x)\|_{\mathrm{op}} \le \varepsilon$。

我们指出，在基于常微分方程的生成模型误差分析文献中，对 $v(t,x)$ 的导数估计精度作出此类假设是标准做法（Li 等人,2024年a, 2025年; Liu 等人, 2025年）。第二项意味着一致散度估计误差存在一个误差界，与第一项结合后，有助于界定离散化误差。我们还强调，在紧致 流形上可获得一致估计误差界（Yarotsky,2017; Mena 等人, 2025年）。

在我们的假设下，可以证明以下关于全变差距离的上界（见定义 A.1）。

定理1。设假设 *1*、*2* 和 *3*成立。定义

$$\mathsf{C}_{Lip} := 3\sqrt{L_t^{score}} L_t^{v,x} L^v + \sqrt{L_t^{score}} L_t^{v,t} + 3L^v L_t^{\mathrm{div},x} + L_t^{\mathrm{div},t} + L_R (L^v)^2 d \tag{5}$$

$$\mathsf{C}_{eps} := \sqrt{2L_t^{score}} + 1 + 2\sqrt{L_t^{score}} L^v + \sqrt{L_t^{score}} L_t^{v,x} + L_t^{\mathrm{div},x} + 2L_R d L^v \tag{6}$$

选取恒定步长 h 使其满足引理*5.2*中的要求，可得

$$\mathrm{TV}\left(\pi_T, \hat{\pi}_T\right) \le h\mathsf{C}_{Lip} + \varepsilon \mathsf{C}_{eps} + \varepsilon^2(\sqrt{L_t^{score}} + L_R d).$$

我们对步长施加了上界（如引理 5.2 中所述），以确保每个欧拉步都处于受控邻域内——特别是在指数映射不会发生“折叠”的单射半径范围内，并确保映射 $F_{t_k,h}(x) = \text{指数映射}_x(h\hat{v}(t_k,x))$ 保持可逆——从而使连续时间插值及其关联的向量场 $\tilde{v}$ 均有明确定义。有关更多细节，请参见第5节中关于 h选择的讨论。从较高层面来看，常数 $\mathsf{C}_{\mathsf{Lip}}$ 汇总了总体流（以及得分函数）的利普希茨正则性模量，这些模量控制着局部欧拉截断误差在随时间传输过程中的放大程度；因此，它决定了全变差界中的 $\mathcal{O}(h)$离散化项。相比之下，$\mathsf{C}_{\mathsf{eps}}$ 汇集了相同的几何因素和依赖于得分函数的稳定性因素，但这些因素与估计扰动 $\hat{v} - v$（及其散度）相关，用于量化一个 ε精确的向量场估计如何通过连续性方程传播；这正是它乘在首要的$\mathcal{O}(\varepsilon)$项上的原因，而剩余的 $\varepsilon^2(\sqrt{L_t^{\mathrm{score}}} + L_R d)$则捕捉了得分正则性与向量场失配之间的高阶相互作用。

注释*1.* 提前停止误差可在*Wasserstein*距离中得到控制（见定义*A.1*）：$W_1(\pi_1, \pi_T) \lesssim 1 - T$。在紧流形上，*TV*界蕴含一个 W_1 界，即$W_1(\pi_T, \hat{\pi}_T) \le \mathit{diam}(M)\, \mathit{TV}(\pi_T, \hat{\pi}_T)$（*Villani*，*2008*，定理*6.15*）。因此，我们可以将*TV*界转换为 W_1 距离中的界，并得到 $W_1(\hat{\pi}_T, \pi_1)$的误差界，同时考虑采样误差和提前停止误差。

现在，我们给出 d维超球面上的显式收敛速率。按照Li 等人(2024年a，2025年）；Liu 等人 (2025年)，我们考虑向量场估计良好的情形（即 $\varepsilon \approx 0$），并报告达到 ε_{target} 接近 π_T所需的采样步数 N。

命题 3.1. 对于某个有限的 $d \ge 4$，令 $M = \mathcal{S}^d$，且 $T < 1$ 为早停时间。假设π_1 具有光滑密度并满足 $0 < m_1 \le p_1(x) \le M_1$。选取 $\pi_0(x)$ 为超球面上的均匀分布。则假设 *2*成立，且假设 *2* 中的常数为多项式阶（见引理 *B.5*）。现在若假设 *3* 成立，为达到 ε_{target} 的全变差距离精度，

- 对于如 (14) 所示的恒定步长，复杂度为 $N = \mathcal{O}(d^2/\varepsilon_{target}(1-T)^2)$。
- 对于如(15)中那样精心设计的步长调度，可以将其改进为 $N = \mathcal{O}(d^2/\varepsilon_{target}(1-T))$。

4 总变差率：Hadamard流形

本节将建立 Hadamard 流形（非紧）上关于全变差距离的误差界，涵盖对称正定流形等应用（Li 等人，2024b； Collas等人，2025年）。在保留假设1中相同曲率条件的同时，我们对假设2 和 3 作如下修改，以自然地处理 Hadamard 流形的非紧性。

假设4 （真实向量场的正则性）。 真实向量场 v 满足 $\forall t \in [0,1)$:

(1) $\mathbb{E}_{x_t}\|\nabla v(t,x_t)\|_{\mathrm{op}}^2 \le (L^v_{,t}{}^{x})^2$ 且 $\mathbb{E}_{x_t}\left\|\frac{d}{dt}v(t,x_t)\right\|^2 \le (L^v_{,}{}^{t})^2$。(2) $\mathbb{E}_{x_t}\left|\frac{d}{dt}\text{散度 } v(t,x_t)\right| \le L^{\mathrm{div}}_{,}{}^{t}$ 且 $\mathbb{E}_{x_t}\|\text{梯度 散度 } v(t,x_t)\|^2 \le (L^{\mathrm{div}}_{t}{}^{x})^2$。(3) $\mathbb{E}_{x_t}\|v(t,x_t)\|^2 \le (L^v_t)^2$。

假设 *5*（学习向量场的正则性）。假设学习向量场 $\hat{v}$满足：*(1)* $\|\hat{v}(t,x)\| \le L^{\hat{v}}_t$，*(2)* $\|\nabla\hat{v}(t,x)\| \le L^{\hat{v}}_{t}{}^{x}$ 且 *(3)* $\|\text{梯度 散度 } \hat{v}(t,x)\| \le L^{\mathrm{div}}\hat{}_{t}{}^{v\,x}$，，其中 $L^{\hat{v}}_t$，$L^{\hat{v}}_{t}{}^{x}$，$L^{\mathrm{div}}\hat{}_{t}{}^{v\,x}$ 与 L^v_t，$L^v_{t}{}^{x}$，$L^{\mathrm{div}}_{t}{}^{x}$ 的阶数相同。

假设 6 （估计误差）。 存在某个较小的 $\varepsilon > 0$，使得以下条件成立 $\forall t \in \{t_k\}_{k=0}^{N-1}$: （*1*）$\mathbb{E}_{x_t}[\|\hat{v}(t_k, X_{t_k}) - v(t_k, X_{t_k})\|^2] \le \varepsilon$，（*2*）$\mathbb{E}_{x_t}[|\text{散度 }\hat{v}(t_k, X_{t_k}) - \text{散度 } v(t_k, X_{t_k})|] \le \varepsilon$。

假设 4、5 和6 的条件与假设 2 和 3起到相同的作用。在非紧流形上，与欧几里得空间类似，假设正则性在期望意义下成立，而非均匀成立，更为自然。与紧情形相比，我们还需要额外施加假设 5，这对于保证采样常微分方程具有良好性质至关重要。要求学习向量场的 Lipschitz 常数与真实向量场的 Lipschitz 常数同阶，纯粹是为了避免给出形式更复杂的界。请注意，对于紧情形，由于我们可以假设逐点正则性，假设5 中的条件可直接由假设 2 和 3推出。我们还指出，对于欧几里得空间中的现有研究，Liu 等人 (2025年) 和Huang等人(2025年) 假设学习向量场具有一致有界的正则性，且估计误差在期望意义下衡量；这些假设与假设5 和6在精神上相似。下面给出我们的收敛速率。

定理 2 （Hadamard 流形上的采样误差）。 设假设 *1*、*4*、*5* 和 *6*均成立。选取步长 h 使其满足引理 *5.2*中的要求，我们有

$$\mathrm{TV}\left(\pi_T, \hat{\pi}_T\right) \le h\mathsf{C}_{\mathit{Lip}} + \varepsilon\mathsf{C}_{\mathit{eps},1},$$

其中 $\mathsf{C}_{\mathit{Lip}}$ 如 (5) 中所示，且 $\mathsf{C}_{\mathit{eps},\,1}$ 定义于 (13) 表示（向量场）估计误差。

现在，我们将结果具体应用于 $M \equiv$ 对称正定矩阵 $(n) := \{X \in \mathbb{R}^{n\times n}: X \succ 0, X^T = X\}$，即赋有仿射不变度量的对称正定矩阵流形$g_X(U,V) = \mathrm{tr}(X^{-1}UX^{-1}V)$。为此，我们计算正则性常数 $\mathsf{C}_{\mathsf{Lip}}$和 $\mathsf{C}_{\mathsf{eps},\,1}$ 的显式阶数（见命题 B.6），进而得到下面的结果。

命题 4.1. 设 $M =$ 对称正定矩阵 (n)，其维数 $d = n(n+1)/2$。我们采用提前停止，即在区间 $[0,T] \subsetneq [0,1]$ 上模拟常微分方程。选择先验分布为$\pi_0(x) \propto \exp(-\frac{n(n+1)d_g(\quad x\,I\,^2,)}{2})$，其为黎曼高斯分布（黎曼高斯分布的中心是任意的，为记号简便，此处设为 $I \in$ 对称正定矩阵 (n)）。我们进一步假设数据分布 π_1 满足

$$\max\left\{\mathbb{E}[d(X_1,I)^2 e^{\lambda_1 d(X_1,I)}], \mathbb{E}[e^{\lambda_1 d(X_1,I)}]\right\} \le M_{\lambda_1}, \quad \textit{where} \quad \lambda_1 = 24\max\{1,\kappa\}, \tag{7}$$

对某个常数 M_{λ_1}。则假设 4 成立，且假设4 中的常数为多项式阶（显式界见命题 B.6 ）。在假设 1、5 和假设6下，（采用恒定步长）在全变差距离下达到 ε_{target} 精度的迭代复杂度为$_3N = \mathcal{O}($ $d^{24}L_R^3 M_{\lambda_1}^2/\varepsilon_{target}(1-T)^3)$。

这里，矩条件（7）发挥的作用类似于欧氏离散化分析中出现的有限矩条件（例如，$\mathbb{E}[\|X_1\|^4] < \infty$ 以及有界支撑等更强的条件）（Liu 等人，2025年；Li 等人，2025年；Zhou 和 Liu，2025年）。指数矩的出现源于曲率畸变：模型函数 $s_{K_{\min}}(r$关于 r 呈线性增长$K_{\min} = 0$ ，但关于 r 呈指数增长 $K_{\min} < 0$ （参见第 5.2.1 节）。尽管 d 的阶为多项式阶，但我们预计，通过更精细的计算和附加假设，还可以进一步改进。

5 证明概要与中间结果

现在，我们概述主要全变差收敛速率界的证明思路。从整体上看，我们将总体流匹配动力学及其数值近似都视为流形上的确定性输运，这些输运由（可能不同的）时变向量场驱动。随机性仅来自初始化，因此我们关注的是这些输运所诱导的时间边缘分布 $(p_t)^{t\in[0,1]}$和 $(q_t)^{t\in[0,1]}$ 。我们的论证分为三个概念性步骤：

1.（输运下的总变差稳定性）推导 ∂_t TV（p_t， q_t）的微分恒等式，并用驱动向量场之间的不匹配来表示（引理 5.1）；2.（欧拉格式的连续时间插值）构造欧拉格式的一个插值，使其本身是 M上的常微分方程，从而可以应用引理 5.1 ；3.（逐项控制）通过证明（i）得分正则性 p_t，（ii）相关向量场的正则性，以及（iii）插值场散度的曲率相关估计，来界定由此得到的右端项。

随后对微分不等式随时间积分，即可得到终端采样误差 TV（X_1， Y_1）的界（或在需要时，得到极限 $t\uparrow 1$ 的界）。

我们首先介绍关键的总变差导数引理，它是（Li 等人，2025年，引理 4.2）的黎曼类比。证明包含以下步骤：（i）通过连续性方程写出 p 和 q 的演化；（ii）对 $\int(p-q)_+$求导并分部积分，但所有微分算子都必须作内蕴解释（梯度、散度和黎曼体积形式）。（10）的第二行通过取绝对值并对内积项应用柯西–施瓦茨不等式得到。

引理 5.1。设 X_t、Y_t 是定义在 M 上的随机过程，满足以下常微分方程：

$$dX_t = v(t, X_t)\,dt, \qquad X_0 \sim p_0 \tag{8}$$
$$dY_t = \tilde{v}(t, Y_t)\,dt, \qquad Y_0 \sim q_0 \tag{9}$$

令 $p(t,x)$ 为 X_t 的分布，令 $q(t,x)$ 为 Y_t 的分布。定义 $\Omega_t = \{x \in M\colon p(t,x) - q(t,x) > 0\}$。那么我们有

$$\begin{aligned}\frac{\partial\,\mathrm{TV}\,(X_t, Y_t)}{\partial t} &= \int_{\Omega_t} p(t,x)\left(\operatorname{div}\left(\tilde{v}(x,t) - v(x,t)\right) + \langle \operatorname{grad}\,\log p(t,x), \tilde{v}(x,t) - v(x,t)\rangle\right) dV_g(x)\\ &\le \mathbb{E}[|\operatorname{div}\left(\tilde{v}(x,t) - v(x,t)\right)|] + \mathbb{E}[\|\operatorname{grad}\,\log p(t,x)\|\|\tilde{v}(x,t) - v(x,t)\|].\end{aligned} \tag{10}$$

5.1 连续时间插值

现在我们说明如何利用引理 5.1 分析欧拉离散化误差。我们将 (8) 视为由（总体）向量场驱动的流匹配常微分方程 $v(t,\cdot)$，并希望 (9) 表示数值方法。欧拉格式的直接插值为

$$y_t \;=\; \mathrm{Exp}_{y_k}\big((t-t_k)\hat{v}(t_k,y_k)\big), \qquad t\in[t_k,t_{k+1}),$$

也就是说，速度在左端点处保持不变。然而，这条曲线并不能直接写成 $dy_t=\tilde{v}(t,y_t)\,dt$ 的形式，其中 M 上存在一个向量场 $\tilde{v}(t,\cdot)$，因为冻结的向量 $\hat{v}(t_k,y_k)$ 属于 $T_{y_k}M$，而常微分方程要求瞬时速度属于 $T_{y_t}M$。

自然的解决办法是沿曲线平行移动冻结的速度。具体而言，对于 $t\in[t_k,t_{k+1})$，我们定义

$$\tilde{v}(t,y_t):=P_{y_k}^{y_t}\,\hat{v}(t_k,y_k), \qquad \text{so that} \qquad dy_t=\tilde{v}(t,y_t)\,dt,$$

其中 $P_{y_k}^{y_t}$ 表示沿（插值后的）轨迹进行平行移动。这确保对所有 t，$\tilde{v}(t,y_t)\in T_{y_t}M$，因此该插值是 M 上一个真正的常微分方程。要应用引理 5.1，我们还需要将 $\tilde{v}$ 定义为 (t,x) 的函数（而不是通过某条特定路径隐式定义）。这要求给定 (t,x) 后，我们能够唯一地恢复在欧拉插值下生成 x 的基点 y_k。等价地，我们需要映射 $y_k\longmapsto$ 指数映射 $_{y_k}\big((t-t_k)\hat{}v(t_k,y_k)\big)$ 在相关步长范围内可逆。

为方便记号，对于 $t\in[0,1)$ 和 $h\geq 0$，我们定义 $F_{t,h}(x):=$ 指数映射$_x(h\hat{}v(t,x))$. 特别地，对于 $t\in[t_k,t_{k+1})$，欧拉插值可写为 $F_{t_k,t-t_k}(x_k)\;=$指数映射 $_{x_k}((t-t_k)\hat{}v(t_k,x_k))$. 下文的引理 5.2 表明，当 h 充分小时，映射 $F_{t_k,h}$ 可逆。假设其可逆，我们可以通过回拉至唯一的原像 $x_k=F_{t_k,t-t_k}^{-1}(x)$，再将固定速度平行移动至 T_xM，从而表示任意点 x 处的插值向量场：

$$\tilde{v}(t,x)=P^{x}_{F^{-1}_{t_k,t-t_k}(x)}\hat{v}(t_k,F^{-1}_{t_k,t-t_k}(x)),$$

其中，对于（9）的一条轨迹 (x_t)，有 $F_{t_k,t-t_k}^{-1}(x_t)=x_{t_k}$。

引理 5.2。 设 M 是满足假设 *1*的单连通黎曼流形。令b为 M上的任意向量场，满足 $\|b(x)\|\leq B$, $\forall x\in M$。假设 $\|\nabla_v b(x)\|\leq L_\nabla\|v\|$。令$R=\mathrm{inj}\,(M)$。为保证 $F_{t_k,t-t_k}$ 可逆，我们要求

$$h<\min\{\frac{R}{B},\frac{1}{4L_\nabla},\sqrt{\frac{3}{4\|b(x)\|^2L_R(2+2L_\nabla\max\{\frac{1}{\sqrt{K_{\min}}},1\})}}\}, \qquad \textit{if} \quad K_{\text{m在}}>0,$$

$$h<\min\{\frac{R}{B},\frac{1}{4L_\nabla},\sqrt{\frac{3}{4\|b(x)\|^2L_R(2\frac{\sinh(\sqrt{-K_{\min}})}{\sqrt{-K_{\min}}}+4\frac{\cosh(\sqrt{-K_{\min}})-1}{-K_{\min}}L_\nabla)}}\}, \qquad \textit{if} \quad K_{\text{m在}}<0,$$

$$h<\min\{\frac{R}{B},\frac{1}{4L_\nabla},\sqrt{\frac{3}{4\|b(x)\|^2L_R(2+hL_\nabla)}}\}, \qquad \textit{if} \quad K_{\text{m在}}=0.$$

可逆性要求准确来说，就是要求单个欧拉步中使用的指数映射始终处于不会“折叠”流形的范围内（例如，不会穿过共轭点或超出单射半径）。如果 $(t-t_k)\hat{}v(t_k,Y_{t_k})$ 过大，映射 $x\mapsto$指数映射 $_x$ $((t-t_k)\hat{}v(t_k,x))$ 可能不再是单射，从而多个原像可能映射到同一个 Y_t。引理 5.2 通过结合 (i) 单射半径控制以及

(ii) 基于曲率比较和协变导数估计的 $F_{t,h}$ 微分界。当 inj $(M) = \infty$（例如，$M = \mathbb{R}^d$ 或 $M =$ 对称正定矩阵 (n)）时，单射半径约束不再起作用，因此步长条件也相应简化。

5.2 离散化分析

尽管（8）和（9）是确定性常微分方程，但随机初始化使得（X_t）和（Y_t）通过其诱导的分布成为随机过程。为隔离离散化误差，我们令它们具有相同的初始分布：$X_0 \sim \pi_0$，以及 $Y_0 \sim \pi_0$.。于是，TV（X_t，Y_t）衡量总体流匹配输运与欧拉插值所诱导输运之间的差异。特别地，在终止时刻（或在极限 $t \uparrow 1$）下，TV（X_1，Y_1）恰好是可归因于时间离散化的采样误差（以及在存在时，使用$\hat{}v$ 而非 v所造成的误差）。

应用引理5.1 ，令 p（t，$\cdot$）为 X_t 的分布，以及 q（t，$\cdot$）为 Y_t的分布，并对时间进行积分，可得

$$\mathrm{TV}\,(X_1, Y_1) \leq \mathrm{TV}\,(X_0, Y_0) + \int_0^1 (\mathbb{E}[|\,\mathrm{div}\,(\tilde{v}(t, X_t) - v(t, X_t))|] +$$

$$\mathbb{E}[\|\text{梯度 对数映射 } p_t(X_t)\|\|\tilde{v}(t,\ X_t) - v(t,\ X_t)\|])dt.$$

由于在共享初始化下全变差（X_0, Y_0）$= 0$，任务归结为控制右端的两个误差项。组织分析的一种便捷方式，是对论证中反复出现的三个量给出界：$\mathbb{E}\big[\|\text{梯度 log } p_t(X_t)\|^2\big]$、$\mathbb{E}\big[\|\tilde{v}(t, X_t) - v(t, X_t)\|^2\big]$，以及$\mathbb{E}\big[|\,\text{散度 }(\tilde{v}(t, X_t) - v(t, X_t))|\big]$。具体而言，(10) 中的第二个期望可通过柯西–施瓦茨不等式解耦：

$$\mathbb{E}[\|\,\mathrm{grad}\,\log p_t(X_t)\|\|\tilde{v}(t, X_t) - v(t, X_t)\|] \leq \sqrt{\mathbb{E}[\|\,\mathrm{grad}\,\log p_t(X_t)\|^2]}\sqrt{\mathbb{E}[\|\tilde{v}(t, X_t) - v(t, X_t)\|^2]}.$$

因此，一旦我们对这三项建立适当的界（对远离 1 的 t 一致成立，并且显式依赖于 t 随 $t \uparrow 1$的变化），便可对时间积分，从而得到定量的全变差收敛速率。

5.2.1 得分正则性与向量场正则性

我们现在总结为控制得分项与向量场失配所需的正则性条件。得分函数界是主要的“密度正则性”成分，而向量场界则确保：(i) 欧拉插值是良定义的；(ii) 失配 $\tilde{v} - v$ 可以在几何适配范数下加以量化。

• **得分正则性**。我们证明存在一个（可能依赖于时间的）常数 $L_{\text{得分函数}t}$，使得对所有 $t \in [0, 1)$，$\mathbb{E}\big[\|$梯度 log $p_t(X_t)\|^2\big] \leq L_{\text{得分函数}t}$ $\cdot$该界允许随着$t \uparrow 1$而变差；这在一般情况下不可避免，并且与总体条件流场在终止时刻附近的行为一致。我们在命题 5.3 中对紧流形、在命题 5.4 中对 Hadamard流形建立了这一结果（二者均在下文陈述）。一个关键点是，这样的 $L_{\text{得分函数}t}$ 的存在依赖于 p_t 具有足够的光滑性/正性；若不具备这一点，梯度 $\log p_t$ 可能不是平方可积的，并且即使对 $t < 1$ 该界也可能失效。

• **紧流形上的正则性**。在紧流形上，曲率有界且距离受均匀控制，但割点的存在可能会妨碍条件密度的光滑性，例如 $p_t(x_1 \mid x_t = x)$，而这些条件密度会出现在总体流场的显式公式中。我们通过利用球面上的显式计算，在附录2 中验证 $\mathcal{S}^d$ 上的假设D。对于更一般的紧几何，要控制割迹处的光滑性要复杂得多，这也是我们将所需假设抽象表述的主要原因。

• Hadamard 流形上的正则性。在Hadamard流形上，不存在割点，因此指数映射和相关条件密度处处光滑，这大大简化了可微性问题。代价是负曲率和无界距离可能会放大沿测地线的导数界。具体而言， $v(t,x)$的导数逐点界通常会涉及因子 Poly (d)、Poly $(1/(1-t))$ 和 Poly $(s_K\min(d(x_0,x_1)))$；其中，对于 $K\min<0$，模型函数 $s_K\min(r) = \sinh(r\sqrt{-K}\min)\sqrt{-K}\min$ 会随 r呈指数增长。因此，我们倾向于寻求期望意义下的正则性，而非关于 x的均匀界。我们在附录 E 中通过结合曲率比较与矩条件 (7)，验证了对称正定矩阵 (n) 的假设4；该矩条件起到欧氏二阶矩界的作用，可补偿曲率导致的增长。

命题 5.3。 设 M 是满足假设 1的紧流形。在假设2下，存在一个有限常数 L_t^{score}，它依赖于 t，使得 $\mathbb{E}\big[$‖梯度 对数映射 $p_t(X_t)\|^2\big] \le L_t^{\text{score}}$， $\forall t\in[0,1)$。

命题 5.4。 设 M 是一个截面曲率满足 $-\kappa^2 = K_{\min}$ ≤截面曲率 ≤ 0 的 *Hadamard* 流形。若先验分布取为 $X_0 \sim e^{-dd(x_0,z)^2}$，，则有

$$\mathbb{E}\big[\|\operatorname{grad}\log p_t(X_t)\|^2\big] \lesssim M\frac{d^2}{(1-t)^2}L_R^2 d^{2\lambda_{score}},$$

where $\lambda_{score} = 6$最大值$\{1,\kappa\}$ 且 $M = \mathbb{E}_{X_1}[e^{\lambda d(x_1,z)}]$ 依赖于数据分布 $n.$

5.2.2 散度项

最后，我们讨论散度贡献 $\mathbb{E}\big[$|散度 $(\tilde{v}(t,\ X_t) - v(t,\ X_t))|\big]$。这是离散化分析中最具几何特性的部分。在欧几里得空间中，欧拉插值只是冻结速度，无需进行平行移动；因此，div $\tilde{v}$ 可以用初等微积分处理。相比之下，在流形上，我们的插值场 $\tilde{v}$ 通过 (i) 平行移动 和 (ii) 隐式逆映射 $F_{t_k}^{-1}$，$_{t-t_k}$定义。这两种操作都会对 $\nabla\tilde{v}$ 产生不可忽略的影响，进而影响 div $\tilde{v}$。

我们的方法是对该表示进行显式求导

$$\tilde{v}(t,x) = P^{x}_{F^{-1}_{t_k,t-t_k}(x)}\,\hat{v}\big(t_k, F^{-1}_{t_k,t-t_k}(x)\big),$$

for $t\in[t_k,t_{k+1})$，沿着插值测地线，在平行正交标架下，并将 div $\tilde{v}(t,\ x) = \sum_{i=1}^{d}\langle\nabla_{E_i(t)}\tilde{v}(t,\ x)$ $E_i(t)\rangle$ 分解为两部分：

1. 在原像点处计算的 $\hat{}v(t_k,\ \cdot)$ 的导数；2. 对平行移动求导产生的曲率畸变。

结合这些估计，并利用假设 1中的曲率界，我们得到 div $\tilde{v}$ 的界，进而得到 div $(\tilde{v}-v)$ 的界，具体见引理 C.4。这种散度控制与上述得分函数/向量场正则性结果相结合，便完成了积分 (10) 并证明所述全变差收敛速率所需的逐项界。

致谢

KB 是 部分研究由美国国家科学基金会（NSF）资助，项目编号为 DMS-2413426 .

参考文献

F. Alimisis、A. Orvieto、G. B´ecigneul 和 A. Lucchi。用于对黎曼优化中的加速进行建模的连续时间视角。载于《国际人工智能与统计会议》，第1297–1307页。PMLR，2020年。（引见于第54页。）V. Bansal、S. Roy、P. Sarkar 和 A. Rinaldo。整流流的 Wasserstein 收敛与直线性。arXiv预印本 arXiv:2410.14949，2024年。（引见于第3页。）J. Benton、G. Deligiannidis 和 A. Doucet。流匹配方法的误差界。《机器学习研究汇刊》，2024年。（引见于第3页。）J. Cheeger、D. G. Ebin 和 D. G. Ebin。《黎曼几何中的比较定理》，第9卷。北荷兰出版公司，阿姆斯特丹，1975年。（引见于第17、18和38页。）R. T. Q. Chen 和 Y. Lipman。一般几何空间上的流匹配。载于《第十二届国际学习表征会议》，2024年。网址：https://openreview.net/forum? id=g7ohDlTITL。（引见于第3页。）A. H. Cheng、A. Lo、K. L. K. Lee、S. Miret 和 A. Aspuru-Guzik。用于矩约束结构解析的 Stiefel流匹配。载于《第十三届国际学习表征会议》，2025年。网址：https://openreview.net/forum?id=84WmbzikPP。（引见于第3页。）X. Cheng、J. Zhang 和 S. Sra。通过朗之万马尔可夫链蒙特卡罗在黎曼流形上进行高效采样。《神经信息处理系统进展》，35：5995–6006，2022年。（引见于第3页。）A. Collas、C. Ju、N. Salvy 和 B. Thirion。通过拉回几何对脑连接矩阵进行黎曼流匹配。载于《第三十九届神经信息处理系统年会》，2025年。网址：https://openreview.net/forum?id=NY3LzmUXl7。（引见于第3和9页。）C. Criscitiello 和 N. Boumal。流形上非凸优化的一种加速一阶方法。《计算数学基础》，23(4)：1433–1509，2023年。（引见于第36页。）V. De Bortoli、E. Mathieu、M. Hutchinson、J. Thornton、Y. W. Teh 和 A. Doucet。基于黎曼分数的生成建模。《神经信息处理系统进展》，35：2406–2422，2022年。（引见于第2和3页。）K. Gatmiry 和 S. S. Vempala。黎曼朗之万算法的收敛性。arXiv预印本 arXiv:2204.10818，2022年。（引见于第3页。）Y. Guan、K. Balasubramanian 和 S. Ma。用于流形高精度采样的黎曼近端采样器。载于《第三十九届神经信息处理系统年会》，2025年。网址：https://openreview.net/forum?id=KxhCJc8BOg。（引见于第3页。）Y. Guan、K. Balasubramanian 和 S. Ma。用于凸域生成建模的重尾先验镜像流匹配。载于《第十四届国际学习表征会议》，2026年。网址：https://openreview.net/forum?id=dZKl7uc0XQ。（引见于第3、6和42页。）H. Hirai、H. Nieuwboer 和 M. Walter。流形上的内点法：理论与应用。载于《2023年IEEE第64届计算机科学基础年会（FOCS）》，第2021–2030页。IEEE，2023年。（引见于第98页。）

C.-W. Huang、M. Aghajohari、J. Bose、P. Panangaden 和 A. C. Courville。黎曼扩散模型。《神经信息处理系统进展》，35:2750–2761，2022年。（第3页引用。）D. Z. Huang、J. Huang 和 Z. Lin。基于分数的生成模型的概率流常微分方程收敛性分析。《IEEE 信息论汇刊》，2025年。（第3页和第9页引用。）S. Kobayashi。《微分几何中的变换群》。Springer，1972年。（第98页引用。）L. Kong 和 M. Tao。李群上的动力学朗之万蒙特卡罗收敛性。载于《第三十七届学习理论年会》，第3011–3063页。PMLR，2024年。（第3页引用。）S. Lang。《微分流形与黎曼流形》，第160卷。Springer Science & Business Media，2012年。（第5页引用。）J. M. Lee。《光滑流形导论》。Springer，纽约，2012年。ISBN 978-1-4419-9982-5。（第45页引用。）J. M. Lee。《黎曼流形导论》，第2卷。Springer，2018。（第18、27、36、51、57、87和98页引用。）M. Lezcano-Casado。依赖曲率的有界几何流形上优化的全局收敛速率。arXiv预印本 arXiv:2008.02517，2020年。（第17、87、89、91、92、93、94和95页引用。）G. Li、Y. Wei、Y. Chi 和 Y. Chen。扩散模型概率流常微分方程的尖锐收敛理论。arXiv预印本 arXiv:2408.02320，2024a。（第4、7和8页引用。）M. Li 和 M. A. Erdogdu。求解半定规划的黎曼朗之万算法。《伯努利》，29(4):3093–3113，2023年。（第3页引用。）R. Li、Q. Di 和 Q. Gu。基于确定性采样器的分数扩散模型统一收敛性分析。载于《第十三届国际学习表征会议》，2025年。网址：https://openreview.net/forum?id=HrdVqFSn1e。（第3、4、7、8和10页引用。）Y. Li、Z. Yu、G. He、Y. Shen、K. Li、X. Sun 和 S. Lin。SPD-DDPM：对称正定空间中的去噪扩散概率模型。载于《AAAI人工智能会议论文集》，第38卷，第13709–13717页，2024b。（第9页引用。）Y. Liu、R. Hu、Y. Chen 和 L. Huang。随机插值基于常微分方程的生成模型有限时间收敛性分析。arXiv预印本 arXiv:2508.07333，2025年。（第4、8、9和10页引用。）A. Lou、M. Xu、A. Farris 和 S. Ermon。黎曼扩散模型的规模化。《神经信息处理系统进展》，36:80291–80305，2023年。（第3页引用。）X. Luo、Z. Wang、Q. Wang、X. Shao、J. Lv、L. Wang、Y. Wang 和 Y. Ma。Crystalflow：用于晶体材料的基于流的生成模型。《自然·通讯》，16(1):9267，2025年。（第3页引用。）J. E. Marsden、T. Ratiu 和 R. Abraham。《流形、张量分析与应用》。应用数学科学。Springer，第3版，2002年。ISBN 0-201-10168-S。（第19和45页引用。）E. Mathieu 和 M. Nickel。黎曼连续归一化流。《神经信息处理系统进展》，33:2503–2515，2020年。（第3和7页引用。）

G. Mena、A. K. Kuchibhotla 和 L. Wasserman。整流流的统计性质。arXiv预印本 arXiv:2511.03193，2025年。（第8页引用。）B. K. Miller、R. T. Q. Chen、A. Sriram 和 B. M. Wood。FlowMM：使用黎曼流匹配生成材料。载于《第41届国际机器学习会议论文集》，第235卷，第35664–35686页，2024年。网址：https://proceedings.mlr.press/v235/ miller24a.html。（第3页引用。）S. Roy、A. Rinaldo 和 P. Sarkar。整流流的低维适应：从扩散与随机定位视角出发的新观点。arXiv预印本 arXiv:2601.15500，2026年。（第4页引用。）Y. Song、J. Sohl-Dickstein、D. P. Kingma、A. Kumar、S. Ermon 和 B. Poole。通过随机微分方程进行基于得分的生成建模。载于《国际学习表征会议》，2021年。网址：https://openreview.net/forum?id=PxTIG12RRHS。（第3页引用。）A. Sriram、B. K. Miller、R. T. Chen 和 B. M. Wood。FlowLLM：以大语言模型为基分布的材料生成流匹配。《神经信息处理系统进展》，37:46025–46046，2024年。（第3和7页引用。）M. Su、J. Y.-C. Hu、S. Pi 和 H. Liu。关于流匹配KL散度。arXiv预印本 arXiv:2511.05480，2025年。（第3页引用。）C. Villani。《最优传输：古今》，第338卷。Springer，2008年。（第8页引用。）Z. Wan、Q. Wang、G. Mishne 和 Y. Wang。通过数据几何与去噪器阐明流匹配常微分方程动力学。载于《第四十二届国际机器学习会议》，2025年。网址：https://openreview.net/forum?id=f5czhqYK3H。（第5页引用。）J. Wu、B. Chen、Y. Zhou、Q. Meng、R. Zhu 和 Z.-M. Ma。黎曼神经测地线插值器。arXiv 预印本 arXiv:2504.15736，2025 年。（第 3 和 19 页引用。） X. Xu 、 Z. Zhang、Y. Nakahira、G. Qu 和 Y. Chi。黎曼扩散模型的多项式收敛。arXiv预印本 arXiv:2601.02499，2026年。（第3页引用。）D. Yarotsky。深度ReLU网络逼近的误差界。《神经网络》，94:103–114，2017年。（第8页引用。）A. Yue、Z. Wang 和 H. Xu。ReQFlow：用于高效、高质量蛋白质骨架生成的整流四元数流。载于《第四十二届国际机器学习会议》，2025年。网址：https://openreview.net/forum?id=f375uEmYDf。（第3和7页引用。）Z. Zhou 和 W. Liu。深度生成建模的流匹配误差分析。载于《第四十二届国际机器学习会议》，2025年。网址：https://openreview. net/forum?id=vES22INUKm。（第3、10和42页引用。）

附录

A 黎曼几何的补充背景

A.1 所用距离

定义A.1 (全变差距离和 W_1 距离)。 设(M, g) 为一个黎曼流形，令 d 表示由 g诱导的测地距离，并设 ρ_1, ρ_2 为 *Borel*σ-代数 $\mathcal{B}(M)$上的概率测度。

- ρ_1 与 ρ_2 之间的全变差距离为

$$\mathrm{TV}\left(\rho_1, \rho_2\right) := \sup_{A \in \mathcal{B}(M)} \left|\rho_1(A) - \rho_2(A)\right|.$$

若 ρ_1 和 ρ_2 相对于黎曼体积测度 dV_g绝对连续，且其密度分别为 $p_1 = \frac{d\rho_1}{dV_g}$ 和 $p_2 = \frac{d\rho_2}{dV_g}$，则

$$\mathrm{TV}\left(\rho_1, \rho_2\right) = \frac{1}{2}\int_M \left|p_1(x) - p_2(x)\right| dV_g(x) = \int_{\{p_1 > p_2\}} \left(p_1(x) - p_2(x)\right) dV_g(x).$$

- *1-Wasserstein*距离（由 d诱导）定义于一阶矩有限的测度上（例如，$\int_M d(x_0, x)\, d\rho_i(x) < \infty$ 对于某些 $x_0 \in M$和 $i \in \{1, 2\}$），其定义为

$$W_1(\rho_1, \rho_2) := \inf_{\gamma \in \Pi(\rho_1, \rho_2)} \int_{M \times M} d(x, y)\, d\gamma(x, y),$$

其中 $\Pi(\rho_1, \rho_2)$ 表示 (ρ_1, ρ_2)的耦合集合，即定义在 $M \times M$上的概率测度 γ，其第一和第二边缘分布分别为 ρ_1 和 ρ_2。

A.2 比较定理

比较定理（Cheeger等人，1975）将在整个附录中频繁使用。例如，我们在附录 C.1中使用了劳赫定理，这是证明引理 5.2的关键步骤；我们也在附录E中频繁使用比较定理来验证假设4。

这里我们简要介绍比较理论，重点阐述其直观含义。设 $k \in \mathbb{R}$为某个常数，并假设有定义在 f、$g \geq 0$ 上的函数 $[0, t]$。若 f和 g 具有相同的初始值 $f(0) = g$(0)，并且进一步假设 $f'(0) \leq g'$(0)。若我们知道，对于二阶常微分方程，$f'' + \kappa f \leq g'' + \kappa g$，那么至少在局部范围内可以得出 $f \leq g$。更多细节可参见例如 Lezcano-Casado (2020，引理 4.8）。这一观察引出了一种更一般的比较原理，它是黎曼几何中比较定理背后的关键思想。例如，有以下定理。

定理3 （命题4.9，出自Lezcano-Casado （2020））。设 M 是一个具有有界截面曲率的黎曼流形，满足 $K_{\min} \leq$ 截面曲率 $\leq K_{\max}$。设 γ： $[0, r] \to M$ 是一条测地线，并设

X, Y是沿着 γ的向量场，且 $X, Y \perp \gamma'$满足以下常微分方程（X'是沿着 γ的 X 的协变导数）：

$$X'' + R(X, \gamma')\gamma' = Y, \qquad X(0) = 0, X'(0) = 0.$$

假设存在一个连续函数 η，它在 $[0, r]$上为 Y：$\|Y\| \leq \eta$ 提供上界。那么，定义为 $\rho'' + K_{\min}\rho = \eta$ 的解的 ρ，$\rho(0) = 0$，$\rho'(0) = 0$，为 X：$\|X\| \leq \rho$ 提供上界。

总而言之，利用常微分方程的比较原理，只要向量场满足某个特定的常微分方程，我们就可以利用曲率信息，为定义在黎曼流形上的某些向量场建立界。关于比较定理的全面讨论，请参见例如Lee (2018) 和 Cheeger等人(1975)。

A.3 黎曼子流形

我们简要介绍 Lee Lee(2018) 中关于黎曼子流形的一些结果，这些结果将用于证明引理 E.15。设 $N \subseteq M$ 是 M的一个黎曼子流形，并带有诱导度量。为避免歧义，我们记 ∇^M 为 M上的联络，记 ∇^N 为 N上的联络。设 X, Y是 N上的向量场。我们可以将它们延拓为 M上的向量场，而协变导数 $(\nabla^M)_X Y$可以分解为

$$(\nabla^M)_X Y = ((\nabla^M)_X Y)^\perp + ((\nabla^M)_X Y)^\parallel.$$

法向分量 $((\nabla^M)_X Y)^\perp$定义了第二基本形式，它是一个从$\mathfrak{X}(N) \times \mathfrak{X}(N)$到 N的法丛截面的映射，形式化定义如下：

$$\mathrm{II}(X, Y) := ((\nabla^M)_X Y)^\perp.$$

高斯公式（Lee，2018，定理 8.2）指出，对于 X，Y是定义在 N 上的向量场，并任意延拓至 M，则

$$(\nabla^M)_X Y = (\nabla^N)_X Y + \mathrm{II}(X, Y).$$

一个密切相关的术语是全测地流形。 N 如果 N 中的每条测地线也是 M中的测地线，就称其为全测地的。此外，根据Lee （2018，命题 8.12），以下条件等价：

- N 是 M的全测地子流形。
- N 的第二基本形式恒等为零。
- N 中的每条测地线也是 M中的测地线。

B 主定理证明

本节安排如下。我们首先在附录 5.1 中证明引理 B.1，该引理给出了一种控制常微分方程模拟过程中全变差距离传播的方法：

$$\frac{\partial \,\mathrm{TV}\,(X_t, Y_t)}{\partial t} = \int_{\Omega_t} p(t,x)\left(\mathrm{div}\,(\tilde{v}(x,t) - v(x,t)) + \langle \mathrm{grad}\,\log p(t,x), \tilde{v}(x,t) - v(x,t)\rangle\right) dV_g(x).$$

请注意，全变差距离的时间导数由两部分组成：

- 我们控制“速度向量”项 $p\left(t,\ x\right)\langle$梯度 对数映射 $p\left(t,\ x\right)\ \tilde{v}\left(x,\ t\right)\ -v\left(x,\ t\right)\rangle$ 详见附录B.2。

- 我们在附录 $p(t,x)\operatorname{div}\left(\tilde{v}(x,t)\ -v(x,t)\right)$ 中控制“散度项” B.3。

此外，回顾一下，$\tilde{v}$ 定义良好依赖于$F_{t\text{的可逆性}}h(x)$: =指数映射$_x(h\hat{}v(t,x))$，我们将在附录 B.4中论证这一点。主要结果（定理 1 和2中的采样误差界）的证明见附录 B.5。最后，我们分别在附录 B.6 和B.7中给出超球面和 SPD 流形上的迭代复杂度证明。

附录 C 包含定理证明所需的辅助结果1 和2。最后，附录 D 和E 证明了证明命题3.1 和4.1所需的正则性结果。

B.1 引理 5.1 的证明

我们从以下结果开始。

定理 4 (输运定理（Marsden 等人，2002，定理 8.1.12））。设 (M,μ) 为一个体积流形，且 X 是 M 上的一个向量场，其流为 F_t。对于定义在 f 上的光滑函数 $\mathcal{F}(M\times\mathbb{R})$，令 $f_t(m)=f(m,t)$。对于任意开集 $U\subseteq M$，有：

$$\frac{d}{dt}\int_{F_t(U)} f_t\mu=\int_{F_t(U)}\left(\frac{\partial f}{\partial t}+\operatorname{div}_\mu(f_tX)\right)\mu.$$

证明。 [引理51] 我们可以写成

$$\operatorname{TV}\left(X_t,Y_t\right)=\int_{\Omega_t} p(t,x)-q(t,x)dV_g(x).$$

根据输运定理，

$$\begin{aligned}
\frac{\partial\operatorname{TV}\left(X_t,Y_t\right)}{\partial t}&=\frac{d}{dt}\int_{\Omega_t} p(t,x)-q(t,x)dV_g(x)\\
&=\int_{\Omega_t}\left(\frac{\partial(p(t,x)-q(t,x))}{\partial t}+\operatorname{div}\left((p(t,x)-q(t,x))X\right)\right)dV_g(x)\\
&=\int_{\Omega_t}\frac{\partial(p(t,x)-q(t,x))}{\partial t}dV_g(x)+\int_{\partial\Omega_t}(p(t,x)-q(t,x))\langle X,n\rangle dV_{\hat{g}}(x)\\
&=\int_{\Omega_t}\frac{\partial(p(t,x)-q(t,x))}{\partial t}dV_g(x)+\int_{\partial\Omega_t}0\times\langle X,n\rangle dV_{\hat{g}}(x)\\
&=\int_{\Omega_t}\frac{\partial(p(t,x)-q(t,x))}{\partial t}dV_g(x).
\end{aligned}$$

下述连续性方程由Wu 等人（2025年，定理 2）证明。为完整起见，我们给出证明。对于任意光滑的测试函数 φ （若流形非紧且无边界，则还需具有有界支撑），我们有

$$\begin{aligned}
\int_M\varphi(x)\frac{\partial}{\partial t}p(t,x)dV_g(x)=\frac{d}{dt}\int_M\varphi(x)p(t,x)dV_g(x)=\frac{d}{dt}\mathbb{E}[\varphi(x_t)]=\mathbb{E}[\nabla\varphi(x_t)\circ\frac{d}{dt}x_t]\\
=\int_M\langle\operatorname{grad}\ \varphi(x),v(x,t)\rangle p(t,x)dV_g(x)
\end{aligned}$$

$$= -\int_M \varphi(x) \operatorname{div}(v(x,t)p(t,x)) dV_g(x).$$

因此我们得出

$$\frac{\partial}{\partial t} p(t,x) = -\operatorname{div}(v(x,t)p(t,x)).$$

因此

$$\begin{aligned}
\frac{\partial \operatorname{TV}(X_t, Y_t)}{\partial t} &= \int_{\Omega_t} \frac{\partial (p(t,x) - q(t,x))}{\partial t} dV_g(x) \\
&= \int_{\Omega_t} -\operatorname{div}(v(x,t)p(t,x)) + \operatorname{div}(\tilde{v}(x,t)q(t,x)) dV_g(x) \\
&= \int_{\partial\Omega_t} -p(t,x)\langle v(x,t), n\rangle + q(t,x)\langle \tilde{v}(x,t), n\rangle dV_{\hat{g}}(x) \\
&= \int_{\partial\Omega_t} p(t,x)\langle \tilde{v}(x,t) - v(x,t), n\rangle dV_{\hat{g}}(x) \\
&= \int_{\Omega_t} \operatorname{div}((\tilde{v}(x,t) - v(x,t))p(t,x)) dV_g(x) \\
&= \int_{\Omega_t} p(t,x) \operatorname{div}(\tilde{v}(x,t) - v(x,t)) + \langle \operatorname{grad} p(t,x), \tilde{v}(x,t) - v(x,t)\rangle dV_g(x) \\
&= \int_{\Omega_t} p(t,x) \left(\operatorname{div}(\tilde{v}(x,t) - v(x,t)) + \langle \operatorname{grad} \log p(t,x), \tilde{v}(x,t) - v(x,t)\rangle\right) dV_g(x),
\end{aligned}$$

其中注意到 $p = q$ 在 $\partial\Omega_t$上，且最后一个等式源于散度乘积法则。注意梯度 $p(t,x) = p(t,x)$梯度对数映射 $p(t,x)$，从而完成证明。 □

B.2 速度向量项

在本节中，我们对速度向量项进行界定。我们指出，引理 B.1 与引理 B.2 本质上相同，只是它们基于不同的假设。

引理 B.1. 在假设*2*和*3*下，我们可以界定

$$\begin{aligned}
&\mathbb{E}[\|\operatorname{grad} \log p(t,x)\| \cdot \|\tilde{v}(x,t) - v(x,t)\|] \\
\leq &\varepsilon\sqrt{2L_t^{score}} + \left(L_t^{score}(2(t-t_k)^2(L_t^{\hat{v},x}L^{\hat{v}} + L^v L_t^{\hat{v},x} + L_t^{v,t} + L^v L_t^{v,x})^2)\right)^{\frac{1}{2}}.
\end{aligned}$$

证明。 利用三角不等式，我们可以写成

$$\begin{aligned}
&\|\tilde{v}(t,X_t) - v(t,X_t)\| \\
\leq &\|P_{X_t}^{X_{t_k}}\tilde{v}(t,X_t) - \hat{v}(t_k,X_{t_k})\| + \|\hat{v}(t_k,X_{t_k}) - v(t_k,X_{t_k})\| + \|v(t_k,X_{t_k}) - P_{X_t}^{X_{t_k}} v(t,X_t)\|.
\end{aligned}$$

记作 $X_{t\to t_k} = F_{t_k}^{-1}{}_{,\ t-t_k}(X_t)$。

对于上面的第一项，我们有

$$\begin{aligned}
&\|P_{X_t}^{X_{t_k}}\tilde{v}(t,X_t) - \hat{v}(t_k,X_{t_k})\| \\
= &\|P_{X_t}^{X_{t_k}} P_{X_{t\to t_k}}^{X_t}\hat{v}(t_k,X_{t\to t_k}) - \hat{v}(t_k,X_{t_k})\| = \|P_{X_{t\to t_k}}^{X_t}\hat{v}(t_k,X_{t\to t_k}) - P_{X_{t_k}}^{X_t}\hat{v}(t_k,X_{t_k})\|
\end{aligned}$$

$$\begin{aligned}
\leq & \|P_{X_{t\to t_k}}^{X_t}\hat{v}(t_k, X_{t\to t_k}) - \hat{v}(t_k, X_t)\| + \|\hat{v}(t_k, X_t) - P_{X_{t_k}}^{X_t}\hat{v}(t_k, X_{t_k})\| \\
\leq & L_t^{\hat{v},x} d(X_{t\to t_k}, X_t) + L_t^{\hat{v},x} d(X_t, X_{t_k}) = L_t^{\hat{v},x}\left(d(F_{t_k,t-t_k}^{-1}(X_t), X_t) + d(X_t, X_{t_k})\right),
\end{aligned}$$

这里我们利用了平行移动保持范数不变这一事实。

注意， X_t 是 X 在时间 t的轨迹。从 $X_{t\to t_k}$出发，我们有

$$\mathrm{Exp}_{X_{t\to t_k}}((t-t_k)\hat{v}(t, X_{t\to t_k})) = X_t.$$

因此

$$d(F_{t_k,t-t_k}^{-1}(X_t), X_t) = (t-t_k)\|\hat{v}(t, X_{t\to t_k})\| \leq (t-t_k)L^{\hat{v}}.$$

此外，我们知道 $d(X_t, X_{t_k})$是常微分方程轨迹的距离，因此

$$d(X_t, X_{t_k}) \leq \int_{t_k}^{t} \|v(s, X_s)\| ds \leq (t-t_k)L^v.$$

现在，对于第二项，根据假设，我们有 $\|\hat{v}(t_k, X_{t_k}) - v(t_k, X_{t_k})\| \leq \varepsilon$.对于第三项，

$$\begin{aligned}
\|v(t_k, X_{t_k}) - P_{X_t}^{X_{t_k}} v(t, X_t)\| =& \|v(t_k, X_{t_k}) - v(t, X_{t_k})\| + \|v(t, X_{t_k}) - P_{X_t}^{X_{t_k}} v(t, X_t)\| \\
\leq & (t-t_k)L_t^{v,t} + d(X_t, X_{t_k})L_t^{v,x} \leq (t-t_k)(L_t^{v,t} + L^v L_t^{v,x}).
\end{aligned}$$

将上述估计合并，我们得到

$$\begin{aligned}
& \|\tilde{v}(t, X_t) - v(t, X_t)\|^2 \\
\leq & \Big((t-t_k)L_t^{\hat{v},x}(L^{\hat{v}} + L^v) + \varepsilon + (t-t_k)(L_t^{v,t} + L^v L_t^{v,x})\Big)^2 \\
\leq & 2\varepsilon^2 + 2(t-t_k)^2(L_t^{\hat{v},x}L^{\hat{v}} + L^v L_t^{\hat{v},x} + L_t^{v,t} + L^v L_t^{v,x})^2.
\end{aligned}$$

因此

$$\begin{aligned}
& \mathbb{E}[\|\operatorname{grad}\log p(t,x)\| \cdot \|\tilde{v}(x,t) - v(x,t)\|] \\
\leq & \mathbb{E}[\|\operatorname{grad}\log p(t,x)\|^2]^{\frac{1}{2}} \mathbb{E}[\|\tilde{v}(x,t) - v(x,t)\|^2]^{\frac{1}{2}} \\
\leq & \Big(L_t^{\text{score}}(2\varepsilon^2 + 2(t-t_k)^2(L_t^{\hat{v},x}L^{\hat{v}} + L^v L_t^{\hat{v},x} + L_t^{v,t} + L^v L_t^{v,x})^2)\Big)^{\frac{1}{2}} \\
\leq & \varepsilon\sqrt{2L_t^{\text{score}}} + \Big(L_t^{\text{score}}(2(t-t_k)^2(L_t^{\hat{v},x}L^{\hat{v}} + L^v L_t^{\hat{v},x} + L_t^{v,t} + L^v L_t^{v,x})^2)\Big)^{\frac{1}{2}},
\end{aligned}$$

其中注意，$\sqrt{\mathbb{E}[A+B]} \leq \sqrt{\mathbb{E}[A]} + \sqrt{\mathbb{E}[B]}$。 □

对于 Hadamard流形，当期望满足正则性条件时，我们有以下结果。请注意，证明策略与前一种情况基本相同。

引理 B.2。 在假设 *4*、*5* 和 *6*下，我们可以得到如下界

$$\begin{aligned}
& \mathbb{E}[\|\operatorname{grad}\log p(t,x)\| \cdot \|\tilde{v}(x,t) - v(x,t)\|] \\
\leq & \varepsilon\sqrt{3L_t^{score}} + \Big(L_t^{score}(6(t-t_k)^2\left((L_t^{\hat{v},x})^2(L_t^{\hat{v}})^2 + (L_t^{\hat{v},x})^2(L_t^v)^2 + (L_t^{v,t})^2 + (L_t^v)^2(L_t^{v,x})^2\right))\Big)^{\frac{1}{2}}.
\end{aligned}$$

证明。 利用三角不等式，我们可以写出

$$\begin{aligned}&\|\tilde{v}(t,X_t)-v(t,X_t)\|\\ \leq&\|P_{X_t}^{X_{t_k}}\tilde{v}(t,X_t)-\hat{v}(t_k,X_{t_k})\|+\|\hat{v}(t_k,X_{t_k})-v(t_k,X_{t_k})\|+\|v(t_k,X_{t_k})-P_{X_t}^{X_{t_k}}v(t,X_t)\|.\end{aligned}$$

因此，根据柯西-施瓦茨不等式，

$$\begin{aligned}&\mathbb{E}[\|\tilde{v}(t,X_t)-v(t,X_t)\|^2]\\ \leq&\mathbb{E}[3\|P_{X_t}^{X_{t_k}}\tilde{v}(t,X_t)-\hat{v}(t_k,X_{t_k})\|^2+3\|\hat{v}(t_k,X_{t_k})-v(t_k,X_{t_k})\|^2+3\|v(t_k,X_{t_k})-P_{X_t}^{X_{t_k}}v(t,X_t)\|^2].\end{aligned}$$

记为 $X_{t\to t_k}=F_{t_k,\,t-t_k}^{-1}(X_t)$。

对于第一项，我们有

$$\begin{aligned}&\mathbb{E}\|P_{X_t}^{X_{t_k}}\tilde{v}(t,X_t)-\hat{v}(t_k,X_{t_k})\|^2\\ =&\mathbb{E}\|P_{X_t}^{X_{t_k}}P_{X_{t\to t_k}}^{X_t}\hat{v}(t_k,X_{t\to t_k})-\hat{v}(t_k,X_{t_k})\|^2=\mathbb{E}\|P_{X_{t\to t_k}}^{X_t}\hat{v}(t_k,X_{t\to t_k})-P_{X_{t_k}}^{X_t}\hat{v}(t_k,X_{t_k})\|^2\\ \leq&2\mathbb{E}[\|P_{X_{t\to t_k}}^{X_t}\hat{v}(t_k,X_{t\to t_k})-\hat{v}(t_k,X_t)\|^2+2\|\hat{v}(t_k,X_t)-P_{X_{t_k}}^{X_t}\hat{v}(t_k,X_{t_k})\|^2]\\ \leq&2(L_t^{\hat{v},x})^2\mathbb{E}[d(X_{t\to t_k},X_t)^2]+2(L_t^{\hat{v},x})^2\mathbb{E}[d(X_t,X_{t_k})^2]\\ =&2(L_t^{\hat{v},x})^2\mathbb{E}\left[d(F_{t_k,t-t_k}^{-1}(X_t),X_t)^2+d(X_t,X_{t_k})^2\right],\end{aligned}$$

其中我们使用了平行移动保持范数不变这一事实。

注意， X_t 是 X 在时间 t的轨迹。从$X_{t\to t_k}$出发，我们有

$$\operatorname{Exp}_{X_{t\to t_k}}((t-t_k)\hat{v}(t,X_{t\to t_k}))=X_t.$$

因此

$$\mathbb{E}[d(F_{t_k,t-t_k}^{-1}(X_t),X_t)^2]=\mathbb{E}[(t-t_k)^2\|\hat{v}(t,X_{t\to t_k})\|^2]\leq(t-t_k)^2(L_t^{\hat{v}})^2.$$

此外，我们知道 $d(X_t,\ X_{t_k})$ 是常微分方程轨迹的距离，因此

$$\begin{aligned}\mathbb{E}[d(X_t,X_{t_k})^2]\leq\mathbb{E}[(\int_{t_k}^t\|v(s,X_s)\|ds)^2]&\leq(t-t_k)\mathbb{E}[\int_{t_k}^t\|v(s,X_s)\|^2ds]\\ &=(t-t_k)\int_{t_k}^t\mathbb{E}[\|v(s,X_s)\|^2]ds\leq(t-t_k)^2(L_t^v)^2.\end{aligned}$$

现在，对于第二项，我们有

$$\mathbb{E}\|\hat{v}(t_k,X_{t_k})-v(t_k,X_{t_k})\|^2\leq\varepsilon^2.$$

最后，对于第三项，通过交换导数和平行移动的顺序，我们得到

$$\|v(t_k,X_{t_k})-P_{X_t}^{X_{t_k}}v(t,X_t)\|\leq\int_{t_k}^t\left\|\frac{D}{ds}P_{X_s}^{X_{t_k}}v(s,X_s)\right\|ds=\int_{t_k}^t\left\|\partial_sv(s,X_s)+\nabla_{\dot{X}_s}v(s,X_s)\right\|ds.$$

因此，

$$\mathbb{E}\|v(t_k,X_{t_k})-P_{X_t}^{X_{t_k}}v(t,X_t)\|^2\leq(t-t_k)\int_{t_k}^t\mathbb{E}\left\|\partial_sv(s,X_s)+\nabla_{\dot{X}_s}v(s,X_s)\right\|^2ds.$$

注意到

$$\left\|\partial_s v(s,X_s)+\nabla_{\dot{X}_s} v(s,X_s)\right\|^2 \lesssim \|\partial_s v(s,X_s)\|^2+\|\nabla v(s,X_s)\|^2\|v(s,X_s)\|^2,$$

因此我们有

$$\begin{aligned}\mathbb{E}\|v(t_k,X_{t_k})-P_{X_t}^{X_{t_k}}v(t,X_t)\|^2 \leq& (t-t_k)\int_{t_k}^{t}\mathbb{E}\left\|\partial_s v(s,X_s)+\nabla_{\dot{X}_s} v(s,X_s)\right\|^2 ds\\ \lesssim& (t-t_k)^2\mathbb{E}[\|\partial_s v(s,X_s)\|^2+\|\nabla v(s,X_s)\|^2\|v(s,X_s)\|^2]\\ \lesssim& (t-t_k)^2((L_t^{v,t})^2+(L_t^{v,x}L_t^v)^2).\end{aligned}$$

综合上述估计，我们得到

$$\begin{aligned}&\mathbb{E}[\|\tilde{v}(t,X_t)-v(t,X_t)\|^2]\\ \leq&\mathbb{E}[3\|P_{X_t}^{X_{t_k}}\tilde{v}(t,X_t)-\hat{v}(t_k,X_{t_k})\|^2+3\|\hat{v}(t_k,X_{t_k})-v(t_k,X_{t_k})\|^2+3\|v(t_k,X_{t_k})-P_{X_t}^{X_{t_k}}v(t,X_t)\|^2]\\ \leq&6(L_t^{\hat{v},x})^2\left((t-t_k)^2(L_t^{\hat{v}})^2+(t-t_k)^2(L_t^v)^2\right)+3\varepsilon^2+6(t-t_k)^2\left((L_t^{v,t})^2+(L_t^v)^2(L_t^{v,x})^2\right)\\ =&6(t-t_k)^2\left((L_t^{\hat{v},x})^2(L_t^{\hat{v}})^2+(L_t^{\hat{v},x})^2(L_t^v)^2+(L_t^{v,t})^2+(L_t^v)^2(L_t^{v,x})^2\right)+3\varepsilon^2.\end{aligned}$$

因此

$$\begin{aligned}&\mathbb{E}[\|\operatorname{grad}\log p(t,x)\|\cdot\|\tilde{v}(x,t)-v(x,t)\|]\\ \leq&\mathbb{E}[\|\operatorname{grad}\log p(t,x)\|^2]^{\frac{1}{2}}\mathbb{E}[\|\tilde{v}(x,t)-v(x,t)\|^2]^{\frac{1}{2}}\\ \leq&\left(L_t^{\text{score}}(6(t-t_k)^2\left((L_t^{\hat{v},x})^2(L_t^{\hat{v}})^2+(L_t^{\hat{v},x})^2(L_t^v)^2+(L_t^{v,t})^2+(L_t^v)^2(L_t^{v,x})^2\right)+3\varepsilon^2)\right)^{\frac{1}{2}}\\ \leq&\varepsilon\sqrt{3L_t^{\text{score}}}+\left(L_t^{\text{score}}(6(t-t_k)^2\left((L_t^{\hat{v},x})^2(L_t^{\hat{v}})^2+(L_t^{\hat{v},x})^2(L_t^v)^2+(L_t^{v,t})^2+(L_t^v)^2(L_t^{v,x})^2\right))\right)^{\frac{1}{2}}.\end{aligned}$$

□

B.3 散度项

在本节中，我们对散度项进行界定。我们指出，引理 B.3和引理 B.4 本质上相同，只是它们所基于的假设不同。

记为 $z=F_{t_k,\,t-t_k}^{-1}(x)$。我们知道 $\operatorname{Exp}_z((t-t_k)\hat{}v(t_k,\ z))=x$。回想一下，我们记作

$$\tilde{v}(t,x)=P_{F_{t_k,t-t_k}^{-1}(x)}^{x}\hat{v}(t_k,F_{t_k,t-t_k}^{-1}(x)),$$

只需简单地,

$$\tilde{v}(x)=P_z^{\operatorname{Exp}_z(h\hat{v}(z))}\hat{v}(z)$$

为简化记号。

引理 B.3. 在假设 *2*和 *3*下，我们有

$$\mathbb{E}_{p(t,x)}[\operatorname{div}(\tilde{v}(x,t)-v(t,X_t))]\leq\varepsilon+(t-t_k)\left(L_t^{\operatorname{div},x}(L^{\hat{v}}+2L^v)+L_t^{\operatorname{div},t}+L_R(L^{\hat{v}})^2d\right).$$

证明. 根据引理C.4，我们有

$$\begin{aligned}
&\mathbb{E}_{p(t,x)}[\operatorname{div}(\tilde{v}(x,t)-v(t,X_t))]\\
\leq&\mathbb{E}_{p(t,x)}[|\operatorname{div}\,\hat{v}(t_k,z)-\operatorname{div}\,v(t,X_t)|]+L_R(t-t_k)d\mathbb{E}_{p(t,x)}[\|\hat{v}(t_k,z)\|^2]\\
\leq&\mathbb{E}_{p(t,x)}[|\operatorname{div}\,\hat{v}(t_k,z)-\operatorname{div}\,v(t_k,z)|]+\mathbb{E}_{p(t,x)}[|\operatorname{div}\,v(t_k,z)-\operatorname{div}\,v(t_k,X_{t_k})|]\\
&+\mathbb{E}_{p(t,x)}[|\operatorname{div}\,v(t_k,X_{t_k})-\operatorname{div}\,v(t,X_t)|]+L_R(t-t_k)(L^{\hat{v}})^2d.
\end{aligned}$$

回顾

$$d(F^{-1}_{t_k,t-t_k}(X_t),X_t)=(t-t_k)\|\hat{v}(t,X_{t\to t_k})\|\leq(t-t_k)L^{\hat{v}},$$
$$d(X_t,X_{t_k})\leq\int_{t_k}^{t}\|v(s,X_s)\|ds\leq(t-t_k)L^{v}.$$

对于第一项，我们有

$$\mathbb{E}_{p(t,x)}[|\operatorname{div}\,\hat{v}(t_k,z)-\operatorname{div}\,v(t_k,z)|]\leq\varepsilon.$$

对于第二项，我们有

$$\begin{aligned}
&\mathbb{E}_{p(t,x)}[|\operatorname{div}\,v(t_k,z)-\operatorname{div}\,v(t_k,X_{t_k})|]\leq L_t^{\operatorname{div},x}\mathbb{E}_{p(t,x)}[d(z,X_{t_k})]\\
\leq&L_t^{\operatorname{div},x}\left(\mathbb{E}[d(F^{-1}_{t_k,t-t_k}(X_t),X_t)+d(X_t,X_{t_k})]\right)\\
\leq&L_t^{\operatorname{div},x}(t-t_k)(L^{\hat{v}}+L^{v}).
\end{aligned}$$

类似地，对于第三项，

$$\begin{aligned}
&\mathbb{E}_{p(t,x)}[|\operatorname{div}\,v(t_k,X_{t_k})-\operatorname{div}\,v(t,X_t)|]\\
\leq&\mathbb{E}_{p(t,x)}[|\operatorname{div}\,v(t_k,X_{t_k})-\operatorname{div}\,v(t,X_{t_k})|]+\mathbb{E}_{p(t,x)}[|\operatorname{div}\,v(t,X_{t_k})-\operatorname{div}\,v(t,X_t)|]\\
\leq&L_t^{\operatorname{div},t}(t-t_k)+L_t^{\operatorname{div},x}d(X_t,X_{t_k})\leq(t-t_k)(L_t^{\operatorname{div},t}+L_t^{\operatorname{div},x}L^{v}).
\end{aligned}$$

将上述估计合并起来，我们得到

$$\begin{aligned}
&\mathbb{E}_{p(t,x)}[\operatorname{div}(\tilde{v}(x,t)-v(t,X_t))]\\
\leq&\mathbb{E}_{p(t,x)}[|\operatorname{div}\,\hat{v}(t_k,z)-\operatorname{div}\,v(t_k,z)|]+\mathbb{E}_{p(t,x)}[|\operatorname{div}\,v(t_k,z)-\operatorname{div}\,v(t_k,X_{t_k})|]\\
&+\mathbb{E}_{p(t,x)}[|\operatorname{div}\,v(t_k,X_{t_k})-\operatorname{div}\,v(t,X_t)|]+L_R(t-t_k)(L^{\hat{v}})^2d\\
\leq&\varepsilon+L_t^{\operatorname{div},x}(t-t_k)(L^{\hat{v}}+L^{v})+(t-t_k)(L_t^{\operatorname{div},t}+L_t^{\operatorname{div},x}L^{v})+L_R(t-t_k)(L^{\hat{v}})^2d\\
\leq&\varepsilon+(t-t_k)\Big(L_t^{\operatorname{div},x}(L^{\hat{v}}+2L^{v})+L_t^{\operatorname{div},t}+L_R(L^{\hat{v}})^2d\Big).
\end{aligned}$$

□

引理 B.4. 在假设 *4*、*5* 和 *6*下，我们有

$$\begin{aligned}
&\mathbb{E}_{p(t,x)}[\operatorname{div}(\tilde{v}(x,t)-v(t,X_t))]\\
\leq&3\varepsilon+(t-t_k)\Big(L_t^{\operatorname{div}\,\hat{v},x}(L_t^{\hat{v}}+L_t^{v})+L_t^{\operatorname{div},x}L_t^{v}+L_t^{\operatorname{div},t}+L_R(L_t^{\hat{v}})^2d\Big).
\end{aligned}$$

证明. 由引理C.4，我们有

$$\begin{aligned}
&\mathbb{E}_{p(t,x)}[\operatorname{div}(\tilde{v}(x,t)-v(t,X_t))]\\
\leq&\mathbb{E}_{p(t,x)}[|\operatorname{div}\hat{v}(t_k,z)-\operatorname{div}v(t,X_t)|]+L_R(t-t_k)d\mathbb{E}_{p(t,x)}[\|\hat{v}(t_k,z)\|^2]\\
\leq&\mathbb{E}_{p(t,x)}[|\operatorname{div}\hat{v}(t_k,z)-\operatorname{div}v(t_k,z)|]+\mathbb{E}_{p(t,x)}[|\operatorname{div}v(t_k,z)-\operatorname{div}v(t_k,X_{t_k})|]\\
&+\mathbb{E}_{p(t,x)}[|\operatorname{div}v(t_k,X_{t_k})-\operatorname{div}v(t,X_t)|]+L_R(t-t_k)(L^{\hat{v}})^2d.
\end{aligned}$$

回顾

$$\begin{aligned}
\mathbb{E}[d(F^{-1}_{t_k,t-t_k}(X_t),X_t)^2]\leq&(t-t_k)^2(L^{\hat{v}}_t)^2,\\
\mathbb{E}[d(X_t,X_{t_k})^2]\leq&(t-t_k)^2(L^{v}_t)^2.
\end{aligned}$$

对于第一项，我们直接得到

$$\mathbb{E}_{p(t,x)}[|\operatorname{div}\hat{v}(t_k,z)-\operatorname{div}v(t_k,z)|]\leq\varepsilon.$$

对于第二项，由三角不等式可得

$$\begin{aligned}
\left|\operatorname{div}v(t_k,z)-\operatorname{div}v(t_k,X_{t_k})\right|\leq&\left|\operatorname{div}v(t_k,z)-\operatorname{div}\hat{v}(t_k,z)\right|\\
&+\left|\operatorname{div}\hat{v}(t_k,z)-\operatorname{div}\hat{v}(t_k,X_{t_k})\right|\\
&+\left|\operatorname{div}\hat{v}(t_k,X_{t_k})-\operatorname{div}v(t_k,X_{t_k})\right|.
\end{aligned}$$

在 $p(t,x)$ 下取期望，并使用假设 6,

$$\mathbb{E}_{p(t,x)}\left[\left|\operatorname{div}v(t_k,z)-\operatorname{div}v(t_k,X_{t_k})\right|\right]\leq2\varepsilon+\mathbb{E}_{p(t,x)}\left[\left|\operatorname{div}\hat{v}(t_k,z)-\operatorname{div}\hat{v}(t_k,X_{t_k})\right|\right].$$

此外，由于 $\hat{v}$ 的逐点空间正则性（等价地，$\|\operatorname{grad}_x\operatorname{div}\hat{v}(t_k,\cdot)\|$ 有逐点界），函数 $\operatorname{div}\hat{v}(t_k,\cdot)$ 是利普希茨的:

$$\left|\operatorname{div}\hat{v}(t_k,z)-\operatorname{div}\hat{v}(t_k,X_{t_k})\right|\leq L_t^{\operatorname{div}\hat{v},x}\,d(z,X_{t_k}).$$

最后，利用 $d(z,X_{t_k})\leq d(z,X_t)+d(X_t,X_{t_k})$ 以及以下界

$$\mathbb{E}[d(z,X_t)]\leq(t-t_k)L^{\hat{v}}_t,\qquad\mathbb{E}[d(X_t,X_{t_k})]\leq(t-t_k)L^{v}_t,$$

我们得到

$$\mathbb{E}_{p(t,x)}\left[\left|\operatorname{div}v(t_k,z)-\operatorname{div}v(t_k,X_{t_k})\right|\right]\leq2\varepsilon+(t-t_k)L_t^{\operatorname{div}\hat{v},x}(L^{\hat{v}}_t+L^{v}_t).$$

对于第三项，根据沿曲线的微积分基本定理，我们有

$$\operatorname{div}v(t,X_t)-\operatorname{div}v(t_k,X_{t_k})=\int_{t_k}^{t}\left(\partial_s\operatorname{div}v(s,X_s)+\langle\operatorname{grad}_x\operatorname{div}v(s,X_s),\dot{X}_s\rangle\right)ds.$$

等价地，

$$\left|\operatorname{div}v(t,X_t)-\operatorname{div}v(t_k,X_{t_k})\right|\leq\int_{t_k}^{t}|\partial_s(\operatorname{div}v)(s,X_s)+\langle\operatorname{grad}_x\operatorname{div}v(s,X_s),v(s,X_s)\rangle|\,ds.$$

因此，我们得到

$$\mathbb{E}_{p(t,x)}\left[\left|\operatorname{div}v(t,X_t)-\operatorname{div}v(t_k,X_{t_k})\right|\right]$$

$$
\begin{aligned}
&\leq \mathbb{E}_{p(t,x)}\Big[\int_{t_k}^{t} |\partial_s(\operatorname{div}\, v)(s,X_s) + \langle \operatorname{grad}_x \operatorname{div}\, v(s,X_s), v(s,X_s)\rangle|\, ds\Big] \\
&\leq \int_{t_k}^{t} \mathbb{E}_{p(t,x)}\Big[\, |\partial_s(\operatorname{div}\, v)(s,X_s)| + |\langle \operatorname{grad}_x \operatorname{div}\, v(s,X_s), v(s,X_s)\rangle| \,\Big] ds \\
&\lesssim (t-t_k)(L_t^{\operatorname{div},t} + L_t^{\operatorname{div},x} L_t^v).
\end{aligned}
$$

综上所述，根据上述估计，我们得到

$$
\begin{aligned}
&\mathbb{E}_{p(t,x)}[\operatorname{div}(\tilde{v}(x,t) - v(t,X_t))] \\
\leq &\mathbb{E}_{p(t,x)}[|\operatorname{div}\, \hat{v}(t_k,z) - \operatorname{div}\, v(t_k,z)|] + \mathbb{E}_{p(t,x)}[|\operatorname{div}\, v(t_k,z) - \operatorname{div}\, v(t_k,X_{t_k})|] \\
&+ \mathbb{E}_{p(t,x)}[|\operatorname{div}\, v(t_k,X_{t_k}) - \operatorname{div}\, v(t,X_t)|] + L_R(t-t_k)(L_t^{\hat{v}})^2 d \\
\leq &3\varepsilon + L_t^{\operatorname{div}\,\hat{v},x}(t-t_k)(L_t^{\hat{v}} + L_t^v) + (t-t_k)(L_t^{\operatorname{div},t} + L_t^{\operatorname{div},x} L_t^v) + L_R(t-t_k)(L_t^{\hat{v}})^2 d \\
\leq &3\varepsilon + (t-t_k)\Big(L_t^{\operatorname{div}\,\hat{v},x}(L_t^{\hat{v}} + L_t^v) + L_t^{\operatorname{div},x} L_t^v + L_t^{\operatorname{div},t} + L_R(L_t^{\hat{v}})^2 d\Big).
\end{aligned}
$$

□

B.4 黎曼流形上的结果

请注意，在引理 5.1 中，我们不要求向量场 v，$\tilde{v}$ 是流匹配向量场。为了应用该引理来分析离散化方案，我们将令 v 为流匹配的真实向量场，而 $\tilde{v}$ 对应于学习得到的向量场。

然而，连续时间常微分方程 $dY_t = \tilde{v}(t, Y_t)dt$ 与欧拉离散化方案之间存在差异。因此，要将该引理应用于离散化（即算法 1 中的实际方法），我们需要定义一个连续时间插值。

定义 F为 $F_{t,h}(x) :=$ 指数映射 $_x(h\hat{}v(t, x))$. 请注意，

$$
F_{t_k,t-t_k}(x_k) = \operatorname{Exp}_{x_k}((t-t_k)\hat{v}(t_k,x_k))
$$

对应于欧拉离散化的连续时间插值。然后，我们可以定义一个插值向量场：我们希望定义

$$
dY_t = P_{Y_{t_k}}^{Y_t}\hat{v}(t_k,Y_{t_k})dt \overset{?}{:=} \tilde{v}(t,Y_t)dt.
$$

这里我们使用问号，以强调尚未证明 $\tilde{v}$ 定义良好。我们希望将右侧写成 $(t, Y_t) = (t, F_{t_k,t-t_k}($ $Y_{t_k}))$ 的函数。

引理 (引理重述5.2)。 设 M 是满足假设 *1* 的单连通黎曼流形。设 b 是 M 上任意一个向量场，满足 $\|b(x)\| \leq B, \forall x \in M$。假设 $\|\nabla_v b(x)\| \leq L_\nabla \|v\|$。令 $R = \operatorname{inj}(M)$。为保证 $F_{t_k,t-t_k}$ 可逆：

1. 若 $K_{\min} > 0$，我们要求

$$
h < \min\{\frac{R}{B}, \frac{1}{4L_\nabla}, \sqrt{\frac{3}{4\|b(x)\|^2 L_R(2 + 2L_\nabla \max\{\frac{1}{\sqrt{K_{\min}}}, 1\})}}\}.
$$

2. 若 $K_{\min} < 0$，则我们要求

$$
h < \min\{\frac{R}{B}, \frac{1}{4L_\nabla}, \sqrt{\frac{3}{4\|b(x)\|^2 L_R(2\frac{\sinh(\sqrt{-K_{\min}})}{\sqrt{-K_{\min}}} + 4\frac{\cosh(\sqrt{-K_{\min}})-1}{-K_{\min}} L_\nabla)}}\}.
$$

3. 若 $K_{\min}=0$，则我们要求

$$h<\min\{\frac{R}{B},\frac{1}{4L_\nabla},\sqrt{\frac{3}{4\|b(x)\|^2L_R(2+hL_\nabla)}}\}.$$

证明。 [引理 5 的证明。2] 为简化记号，我们用 h 表示时间步长，用 b 表示映射 F的学习向量场。也就是说，对于 $h\in\mathbb{R}$和 $b\in\mathfrak{X}(M)$，我们记作

$$F_h:M\to M,\qquad F_h(x):=\operatorname{Exp}_x\big(hb(x)\big).$$

我们首先计算 F 的导数。令 $x\in M$ 和 $v\in T_xM$。注意，$dF_h(x)$ 可以视为一个算子，它将 $v\in T_xM$ 映射到 $dF_h(x)v\in T_{F_h(x)}M$。我们计算 $dF_h(x)v$。令 $c(s)$ 为一条光滑曲线，使得 $c(0)=x$ 且 $c'(0)=v$。

定义如下的测地线变分：

$$\Lambda:(-\epsilon,\epsilon)\times[0,1]\to M,\qquad \Lambda(s,t):=\operatorname{Exp}_{c(s)}\big(thb(c(s))\big).$$

对于每个固定的 s，t 方向是从 $c(s)$ 出发、初始速度为 $hb(c(s))$ 的测地线。特别地，对于每个 s，我们有

$$F_h(b(c(s)))=\operatorname{Exp}_{c(s)}(hb(c(s)))=\Lambda(s,1).$$

现在定义一条沿中心测地线 $\gamma(t)$ 的向量场 $J_v:=\Lambda(0,t)=$指数映射$_x\big(thb(x)\big)$ by

$$J_v(t):=\partial_s\Lambda(0,t).$$

那么，根据构造（将左侧视为沿由v诱导的曲线的方向导数），

$$dF_h(x)(v)=\frac{d}{ds}F_h(b(c(s)))\Big|_{s=0}=\partial_s\Lambda(0,1)=J_v(1).$$

这表示 F_h 在 x 处作用于 v 的微分，等于沿测地线 γ的变分场J_v 在 $t=1$ 处的取值。注意，J_v 是雅可比场，因此满足雅可比方程，参见 Lee（2018年，定理 10.1）。初始条件为 $J_v(0)=v$，且 $D_tJ_v(0)=D_t\partial_s\Lambda(0,0)=D_s\partial_t\Lambda(0,0)=h\nabla_vb(x)=:\omega$。

现在我们分析保证 F可逆的条件。定义

$$Y(t)=P^{\gamma(0)}_{\gamma(t)}J_v(t).$$

为证明 $dF_h(x)$可逆，只需证明 $\inf_v\frac{\|dF_h(x)v\|}{\|v\|}>0$，等价地，

$$\|dF_h(x)v\|=\|P^{\gamma(0)}_{\gamma(1)}J_v(1)\|=\|Y(1)\|\geq C\|v\|>0,\forall v\neq 0$$

对某个常数 $C>0$成立。

应用引理 C.5 结合 $c=\gamma$ 与 $Y=J_v$，我们得到

$$Y'(t)=\frac{d}{dt}\Big(P^{\gamma(0)}_{\gamma(t)}J_v(t)\Big)=P^{\gamma(0)}_{\gamma(t)}D_tJ_v(t).$$

再求导一次，并用同一引理结合 $Y = D_t J_v$，我们得 t

$$Y''(t) = \frac{d}{dt}\Big(P_{\gamma(t)}^{\gamma(0)} D_t J_v(t)\Big) = P_{\gamma(t)}^{\gamma(0)} D_t^2 J_v(t).$$

应用雅可比方程 $D_t^2 J_v(t) + R(J_v(t)\ \gamma'(t))\gamma'(t) = 0$，我们得到

$$Y''(t) = P_{\gamma(t)}^{\gamma(0)} D_t^2 J_v(t) = -P_{\gamma(t)}^{\gamma(0)} \left(R(J_v(t), \gamma'(t))\gamma'(t)\right).$$

利用 $J_v(t) = P_{\gamma\ 0}^{\gamma(t)}{}() Y(t)$，我们重写曲率项并得到

$$Y''(t) + P_{\gamma(t)}^{\gamma(0)}\Big(R\big(P_{\gamma(0)}^{\gamma(t)} Y(t), \gamma'(t)\big)\gamma'(t)\Big) = 0.$$

现在，我们逐项对 $Y(t)$ 应用带积分余项的泰勒定理 .

$$\begin{aligned} Y(t) &= Y(0) + tY'(0) + \int_0^t (t-s)Y''(s)ds \\ &= v + th\,\nabla_v b(x) - \int_0^t (t-s) P_{\gamma(s)}^{\gamma(0)}\Big(R\big(P_{\gamma(0)}^{\gamma(s)} Y(s), \gamma'(s)\big)\gamma'(s)\Big)ds. \end{aligned}$$

特别地，对于 $t = 1$,

$$Y(1) - v - h\nabla_v b(x) = -\int_0^1 (1-s) P_{\gamma(s)}^{\gamma(0)}\Big(R\big(P_{\gamma(0)}^{\gamma(s)} Y(s), \gamma'(s)\big)\gamma'(s)\Big)ds.$$

由于我们假设 $\|R(u,v)w\| \le L_R\|u\|\|v\|\|w\|$,

$$\|R\big(P_{\gamma(0)}^{\gamma(t)} Y(t), \gamma'(t)\big)\gamma'(t)\| \le L_R\|\gamma'(t)\|^2\|Y(t)\|.$$

取范数后，我们得到

$$\begin{aligned} \|Y(1) - v - h\nabla_v b(x)\| &\le \int_0^1 (1-s)L_R\|\gamma'(s)\|^2\|Y(s)\|ds \\ &\le h^2\|b(x)\|^2 L_R \int_0^1 \|Y(s)\|ds. \end{aligned}$$

其中注意 $\|\gamma'(s)\| = h\|b(x)\|$。我们指出，只需给 $\int_0^1 \|Y(s)\|ds$ 一个上界。然后，我们只需应用三角不等式，即 $\|Y(1)\| \ge \|v + h\nabla_v b(x)\| - \|Y(1) - v - h\nabla_v b(x)\|$。

现在我们通过比较理论来给 $\int_0^1 \|Y(s)\|ds$ 设上界。由于涉及比较理论的计算较为复杂，我们将其归纳为一个独立引理，见引理 C.2。

根据引理 C.2，当 $K_{\min} > 0$ 时，我们有

$$\begin{aligned} \int_0^1 \|Y(t)\|dt \le& \|v^{\|}\| + \|\omega^{\|}\| + \frac{\|\omega^{\perp}\|}{\sqrt{K_{\min}}} + \|v^{\perp}\| \le 2\|v\| + 2h\|\nabla_v b(x)\| \max\{\frac{1}{\sqrt{K_{\min}}}, 1\} \\ \le& \|v\|(2 + 2hL_\nabla \max\{\frac{1}{\sqrt{K_{\min}}}, 1\}) =: C_+, \end{aligned}$$

其中请注意，我们对 v, ω进行了正交分解。

$$\|Y(1) - v - h\,\nabla_v b(x)\| \le h^2\|b(x)\|^2 L_R C_+.$$

根据三角不等式,

$$\begin{aligned}\|Y(1)\| \geq & \|v + h\nabla_v b(x)\| - \|Y(1) - v - h\nabla_v b(x)\| \\ \geq & \|v\| - h\|\nabla_v b(x)\| - h^2\|b(x)\|^2 L_R C_+ \\ \geq & \|v\| \left(1 - hL_\nabla - h^2\|b(x)\|^2 L_R(2 + 2hL_\nabla \max\{\frac{1}{\sqrt{K_{\min}}}, 1\})\right).\end{aligned}$$

显然需要 $h < 1$，因此我们只需找到满足以下条件的 h：

$$\begin{aligned}1 - hL_\nabla > & h^2\|b(x)\|^2 L_R(2 + 2L_\nabla \max\{\frac{1}{\sqrt{K_{\min}}}, 1\}) \\ > & h^2\|b(x)\|^2 L_R(2 + 2hL_\nabla \max\{\frac{1}{\sqrt{K_{\min}}}, 1\}).\end{aligned}$$

若 $h \leq \frac{1}{4L_\nabla}$，则 $1 - hL_\nabla \geq \frac{3}{4}$。数值 $h \leq \sqrt{\frac{3}{4\|b\ x\ \|^2 L_R\ 2+2L_\nabla 0\ (\text{最大值}\{\ \ \frac{1}{\sqrt{K_{\min}}},1\})}}$ 已足够。

当 $K_{\min} < 0$ 时，我们有

$$\begin{aligned}\int_0^1 \|Y(t)\|dt \leq & \|v^\parallel\| + \|\omega^\parallel\| + \frac{\cosh(\sqrt{-K_{\min}}) - 1}{-K_{\min}}\|\omega^\perp\| + \frac{\sinh(\sqrt{-K_{\min}})}{\sqrt{-K_{\min}}}\|v^\perp\| \\ \leq & \|v\| \left(2\frac{\sinh(\sqrt{-K_{\min}})}{\sqrt{-K_{\min}}} + 4\frac{\cosh(\sqrt{-K_{\min}}) - 1}{-K_{\min}} hL_\nabla\right) =: C_-.\end{aligned}$$

于是我们有

$$\|Y(1) - v - h\,\nabla_v b(x)\| \leq h^2\|b(x)\|^2 L_R C_-.$$

根据三角不等式,

$$\begin{aligned}\|Y(1)\| \geq & \|v + h\nabla_v b(x)\| - \|Y(1) - v - h\nabla_v b(x)\| \\ \geq & \|v\| - h\|\nabla_v b(x)\| - h^2\|b(x)\|^2 L_R C_- \\ \geq & \|v\| \left(1 - hL_\nabla - h^2\|b(x)\|^2 L_R(2\frac{\sinh(\sqrt{-K_{\min}})}{\sqrt{-K_{\min}}} + 4\frac{\cosh(\sqrt{-K_{\min}}) - 1}{-K_{\min}} hL_\nabla)\right).\end{aligned}$$

显然需要 $h < 1$，因此我们只需找到满足条件的 h：

$$\begin{aligned}1 - hL_\nabla > & h^2\|b(x)\|^2 L_R(2\frac{\sinh(\sqrt{-K_{\min}})}{\sqrt{-K_{\min}}} + 4\frac{\cosh(\sqrt{-K_{\min}}) - 1}{-K_{\min}} L_\nabla) \\ > & h^2\|b(x)\|^2 L_R(2\frac{\sinh(\sqrt{-K_{\min}})}{\sqrt{-K_{\min}}} + 4\frac{\cosh(\sqrt{-K_{\min}}) - 1}{-K_{\min}} hL_\nabla).\end{aligned}$$

若 $h \leq \frac{1}{4L_\nabla}$，则 $1 - hL_\nabla \geq \frac{3}{4}$。值 $h \leq \sqrt{\frac{3}{4\|b\ x\ \|^2 L_R\ 2\ \ ()(\sinh(\sqrt{-K_{\min}}) +4^{\cosh(\sqrt{-K_{\min}})-1}_{-K_{\min}} \min L_\nabla)}}$ 即可。

当 $K_{\min} = 0$ 时，我们有

$$\begin{aligned}\int_0^1 \|Y(t)\|dt \leq & \|v^\parallel\| + \frac{1}{2}\|\omega^\parallel\| + \frac{1}{2}\|\omega^\perp\| + \|v^\perp\| \\ \leq & \|v\|\,(2 + hL_\nabla) =: C_=.\end{aligned}$$

于是我们有

$$\|Y(1) - v - h\,\nabla_v b(x)\| \le h^2 \|b(x)\|^2 L_R C_=.$$

根据三角不等式,

$$\begin{aligned}\|Y(1)\| &\ge \|v + h\nabla_v b(x)\| - \|Y(1) - v - h\nabla_v b(x)\| \\ &\ge \|v\| - h\|\nabla_v b(x)\| - h^2\|b(x)\|^2 L_R C_= \\ &\ge \|v\| \left(1 - hL_\nabla - h^2\|b(x)\|^2 L_R(2 + hL_\nabla)\right).\end{aligned}$$

显然需要 $h < 1$，因此我们只需找到 h 使得

$$1 - hL_\nabla > h^2\|b(x)\|^2 L_R(2 + L_\nabla) > h^2\|b(x)\|^2 L_R(2 + hL_\nabla).$$

如果 $h \le \frac{1}{4L_\nabla}$，那么 $1 - hL_\nabla \ge \frac{3}{4}$。取值 $h \le \sqrt{\frac{3}{4\|b\,x\,\|^2 L_R()(2+hL_\nabla)}}$ 即可。 □

B.5 主定理

现在我们证明主定理。我们注意到，定理 1 和2 的证明本质上是相同的。

证明。 [定理 1] 根据引理 5.1，我们有

$$\begin{aligned}&\mathrm{TV}\left(\pi_T, \hat{\pi}_T\right) = \mathrm{TV}\left(\pi_{t_N}, \hat{\pi}_{t_N}\right) \\ \le &\mathrm{TV}\left(\pi_{t_0}, \hat{\pi}_{t_0}\right) + \int_{t_0}^{t_N} \mathbb{E}[\langle \mathrm{grad}\, \log p(t,x), \tilde{v}(x,t) - v(x,t)\rangle]dt + \int_{t_0}^{t_N} \mathbb{E}[|\,\mathrm{div}\,(\tilde{v}(x,t) - v(x,t))|]dt \\ \le &\mathrm{TV}\left(\pi_{t_0}, \hat{\pi}_{t_0}\right) + \int_{t_0}^{t_N} \mathbb{E}[\|\,\mathrm{grad}\, \log p(t,x)\| \cdot \|\tilde{v}(x,t) - v(x,t)\|]dt + \int_{t_0}^{t_N} \mathbb{E}[|\,\mathrm{div}\,(\tilde{v}(x,t) - v(x,t))|]dt,\end{aligned}$$

其中 $\pi_t, \hat{}\pi_t$ 分别表示 X_t, Y_t 的分布。

注意到

$$\int_{t_k}^{t_{k+1}} (t - t_k)dt = (\frac{1}{2}t^2 - tt_k)|_{t_k}^{t_{k+1}} = \frac{1}{2}t_{k+1}^2 - \frac{1}{2}t_k^2 - t_k(t_{k+1} - t_k) = \frac{1}{2}(t_{k+1} - t_k)^2. \tag{11}$$

因此，采用恒定步长进行离散化，我们得到

$$\begin{aligned}&\mathrm{TV}\left(\pi_T, \hat{\pi}_T\right) = \mathrm{TV}\left(X_{t_N}, Y_{t_N}\right) \\ \le &\mathrm{TV}\left(X_{t_0}, Y_{t_0}\right) + \int_{t_0}^{t_N} \mathbb{E}[\|\,\mathrm{grad}\, \log p(t,x)\| \cdot \|\tilde{v}(x,t) - v(x,t)\|]dt + \int_{t_0}^{t_N} \mathbb{E}[|\,\mathrm{div}\,(\tilde{v}(x,t) - v(x,t))|]dt \\ \le &\mathrm{TV}\left(X_{t_0}, Y_{t_0}\right) + \frac{1}{2}Nh^2\Big(\sqrt{2L_t^{\mathrm{score}}}(L_t^{\hat{v},x}L^{\hat{v}} + L^v L_t^{\hat{v},x} + L_t^{v,t} + L^v L_t^{v,x}) \\ &+ L_t^{\mathrm{div},x}(L^{\hat{v}} + 2L^v) + L_t^{\mathrm{div},t} + L_R(L^{\hat{v}})^2 d\Big) + Nh(\varepsilon\sqrt{2L_t^{\mathrm{score}}} + \varepsilon) \\ \le &\mathrm{TV}\left(X_{t_0}, Y_{t_0}\right) + h\Big(\sqrt{L_t^{\mathrm{score}}}(L_t^{\hat{v},x}L^{\hat{v}} + L^v L_t^{\hat{v},x} + L_t^{v,t} + L^v L_t^{v,x}) \\ &+ L_t^{\mathrm{div},x}(L^{\hat{v}} + 2L^v) + L_t^{\mathrm{div},t} + L_R(L^{\hat{v}})^2 d\Big) + (\varepsilon\sqrt{2L_t^{\mathrm{score}}} + \varepsilon). \\ = &\mathrm{TV}\left(X_{t_0}, Y_{t_0}\right) + h\mathsf{C}_{\mathsf{Lip}} + \varepsilon\mathsf{C}_{\mathsf{eps}},\end{aligned}$$

其中注意到 $Nh < 1$，并且我们计算

$$\begin{aligned}
&\sqrt{L_t^{\text{score}}}((L_t^{v,x}+\varepsilon)(L^v+\varepsilon)+L^v(L_t^{v,x}+\varepsilon)+L_t^{v,t}+L^vL_t^{v,x})\\
&+L_t^{\text{div},x}(3L^v+\varepsilon)+L_t^{\text{div},t}+L_R(L^v+\varepsilon)^2d\\
=&\varepsilon^2(\sqrt{L_t^{\text{score}}}+L_Rd)+\varepsilon(2\sqrt{L_t^{\text{score}}}L^v+\sqrt{L_t^{\text{score}}}L_t^{v,x}+L_t^{\text{div},x}+2L_RdL^v)\\
&+3\sqrt{L_t^{\text{score}}}L_t^{v,x}L^v+\sqrt{L_t^{\text{score}}}L_t^{v,t}+3L^vL_t^{\text{div},x}+L_t^{\text{div},t}+L_R(L^v)^2d,
\end{aligned}$$

并记

$$\begin{aligned}
\mathsf{C}_{\mathsf{Lip}} &\coloneqq 3\sqrt{L_t^{\text{score}}}L_t^{v,x}L^v+\sqrt{L_t^{\text{score}}}L_t^{v,t}+3L^vL_t^{\text{div},x}+L_t^{\text{div},t}+L_R(L^v)^2d,\\
\mathsf{C}_{\mathsf{eps}} &\coloneqq \sqrt{2L_t^{\text{score}}}+1+2\sqrt{L_t^{\text{score}}}L^v+\sqrt{L_t^{\text{score}}}L_t^{v,x}+L_t^{\text{div},x}+2L_RdL^v+\varepsilon(\sqrt{L_t^{\text{score}}}+L_Rd).
\end{aligned}$$

□

证明。 [定理的证明 2] 我们有

$$\begin{aligned}
&\mathbb{E}[\|\operatorname{grad}\log p(t,x)\|\cdot\|\tilde{v}(x,t)-v(x,t)\|]\\
\le&\varepsilon\sqrt{3L_t^{\text{score}}}+(t-t_k)\Big(6L_t^{\text{score}}\left((L_t^{\hat{v},x})^2(L_t^{\hat{v}})^2+(L_t^{\hat{v},x})^2(L_t^{v})^2+(L_t^{v,t})^2+(L_t^{v})^2(L_t^{v,x})^2\right)\Big)^{\frac{1}{2}},
\end{aligned}$$

and

$$\mathbb{E}_{p(t,x)}[\operatorname{div}(\tilde{v}(x,t)-v(t,X_t))]\le 3\varepsilon+(t-t_k)\Big(L_t^{\text{div }\hat{v},x}(L_t^{\hat{v}}+L_t^{v})+L_t^{\text{div},x}L_t^{v}+L_t^{\text{div},t}+L_R(L_t^{\hat{v}})^2d\Big).$$

另外回忆

$$\begin{aligned}
&\operatorname{TV}(\pi_T,\hat{\pi}_T)=\operatorname{TV}(\pi_{t_N},\hat{\pi}_{t_N})\\
\le&\operatorname{TV}(\pi_{t_0},\hat{\pi}_{t_0})+\int_{t_0}^{t_N}\mathbb{E}[\langle\operatorname{grad}\log p(t,x),\tilde{v}(x,t)-v(x,t)\rangle]dt+\int_{t_0}^{t_N}\mathbb{E}[|\operatorname{div}(\tilde{v}(x,t)-v(x,t))|]dt\\
\le&\operatorname{TV}(\pi_{t_0},\hat{\pi}_{t_0})+\int_{t_0}^{t_N}\mathbb{E}[\|\operatorname{grad}\log p(t,x)\|\cdot\|\tilde{v}(x,t)-v(x,t)\|]dt+\int_{t_0}^{t_N}\mathbb{E}[|\operatorname{div}(\tilde{v}(x,t)-v(x,t))|]dt.
\end{aligned}$$

Notice that

$$\int_{t_k}^{t_{k+1}}(t-t_k)dt=(\frac{1}{2}t^2-tt_k)|_{t_k}^{t_{k+1}}=\frac{1}{2}t_{k+1}^2-\frac{1}{2}t_k^2-t_k(t_{k+1}-t_k)=\frac{1}{2}(t_{k+1}-t_k)^2. \tag{12}$$

因此，使用恒定步长进行离散化，我们得到

$$\begin{aligned}
&\operatorname{TV}(\pi_T,\hat{\pi}_T)\\
\le&\operatorname{TV}(X_{t_0},Y_{t_0})+\int_{t_0}^{t_N}\mathbb{E}[\|\operatorname{grad}\log p(t,x)\|\cdot\|\tilde{v}(x,t)-v(x,t)\|]dt+\int_{t_0}^{t_N}\mathbb{E}[|\operatorname{div}(\tilde{v}(x,t)-v(x,t))|]dt\\
\le&\operatorname{TV}(X_{t_0},Y_{t_0})+\frac{1}{2}Nh^2\Big(\sqrt{6L_t^{\text{score}}}\left((L_t^{\hat{v},x})^2(L_t^{\hat{v}})^2+(L_t^{\hat{v},x})^2(L_t^{v})^2+(L_t^{v,t})^2+(L_t^{v})^2(L_t^{v,x})^2\right)\\
&+\Big(L_t^{\text{div }\hat{v},x}(L_t^{\hat{v}}+L_t^{v})+L_t^{\text{div},x}L_t^{v}+L_t^{\text{div},t}+L_R(L_t^{\hat{v}})^2d\Big)\Big)+Nh(\varepsilon\sqrt{2L_t^{\text{score}}}+3\varepsilon) \qquad (13)\\
\le&\operatorname{TV}(X_{t_0},Y_{t_0})+h\mathsf{C}_{\mathsf{Lip}}+\varepsilon\mathsf{C}_{\mathsf{eps},\,1},
\end{aligned}$$

其中注意到 $Nh < 1$。取 $t_0 = 0, t_N = T < 1$，我们得到

$$\operatorname{TV}(\pi_T,\hat{\pi}_T)\le h\mathsf{C}_{\mathsf{Lip}}+\varepsilon\mathsf{C}_{\mathsf{eps},\,1}.$$

□

B.6 示例：超球面

对于紧流形，在一致估计误差（假设 3）下，我们可以建立 $\hat{v}$的正则性。具体而言，在假设 3 和假设 2下，由于 $\|\nabla\hat{v}(x)\|_{\mathrm{op}} = \|\nabla\hat{v}(x) - \nabla v(x) + \nabla v(x)\|_{\mathrm{op}}$，我们有

1. $\|\hat{v}(t,\ x)\| \le L^{\hat{v}} = L^v + \varepsilon$。
2. $\hat{v}(t,\ x)$ 关于 x 变量是利普希茨的，常数为 $L^{v,\ t}{}^{x} = L^{v,\ t}{}^{x} + \varepsilon$。

引理 B.5。 在 $\mathcal{S}^d$，假设 2 以如下常数成立。

1. $L_t^{v,x} = \frac{12\pi M_1(d-1)}{m_1(1-t)}$, $L_t^{\hat{v},x} = \varepsilon + \frac{12\pi M_1(d-1)}{m_1(1-t)}$.
2. $L_t^{v,t} := \frac{8\pi^2 d}{1-t}\frac{M_1}{m_1}$.
3. $L_t^{\mathrm{div},x} := \frac{128\pi(d-1)^2}{(1-t)^3}\frac{M_1}{m_1}$.
4. $L_t^{\mathrm{div},t} := \frac{128\pi^2(d-1)^2}{(1-t)^3}\frac{M_1}{m_1}$.
5. $L_t^{score} := \frac{8(d-1)^2}{(1-t)^2}\frac{M_1}{m_1}$.
6. $L^v = \pi$.
7. $L_R = 1$.

由于证明较为技术性，我们将其推迟到附录D。

现在我们证明命题 3.1。 **证明**。 [命题 3.1] 的证明。我们首先回顾引理5.2所施加于 h 的额外条件：

$$h < \min\{\frac{R}{\|b(x)\|}, \frac{1}{4L_\nabla}, \sqrt{\frac{3}{4\|b(x)\|^2 L_R(2 + 2L_\nabla \max\{\frac{1}{\sqrt{K_{\min}}}, 1\})}}\}.$$

由于 $\|\hat{v}(t,x)\| \le L^{\hat{v}} = L^v + \varepsilon$，我们知道 $\|b(x)\|$ 的阶为常数，因此 $\|b^{R}{}_{x()}\|$ 的阶为常数。因此，该条件由项 $\frac{1}{L_\nabla}$主导，而在我们的情形下，该项恰好是 $L^{\frac{1}{\hat{v}}}{}_{,t}{}^{x}$。将常数代入引理 B.5，可得

$$\begin{aligned}
&\mathbb{E}[\|\operatorname{grad}\log p(t,x)\| \cdot \|\tilde{v}(x,t) - v(x,t)\|] \\
\le&\varepsilon\sqrt{2\frac{8(d-1)^2}{(1-t)^2}\frac{M_1}{m_1} + (t-t_k)\Big(\frac{8(d-1)^2}{(1-t)^2}\frac{M_1}{m_1}(2((\varepsilon + \frac{12\pi M_1(d-1)}{m_1(1-t)})(\pi+\varepsilon) \\
&+ \pi(\varepsilon + \frac{12\pi M_1(d-1)}{m_1(1-t)}) + \frac{8\pi^2 d}{1-t}\frac{M_1}{m_1} + \pi\frac{12\pi M_1(d-1)}{m_1(1-t)})^2)\Big)^{\frac{1}{2}}} \\
\le&4\varepsilon\frac{d-1}{1-t}\sqrt{\frac{M_1}{m_1}} + (t-t_k)4\frac{d-1}{1-t}\sqrt{\frac{M_1}{m_1}}\big((\varepsilon + \frac{12\pi M_1(d-1)}{m_1(1-t)})(\pi+\varepsilon) \\
&+ \pi(\varepsilon + \frac{12\pi M_1(d-1)}{m_1(1-t)}) + \frac{8\pi^2 d}{1-t}\frac{M_1}{m_1} + \pi\frac{12\pi M_1(d-1)}{m_1(1-t)}\big) \\
\lesssim&\varepsilon\frac{d-1}{1-t}\Big(\frac{M_1}{m_1}\Big)^{\frac{1}{2}} + (t-t_k)\frac{(d-1)^2}{(1-t)^2}\Big(\frac{M_1}{m_1}\Big)^{\frac{3}{2}}.
\end{aligned}$$

类似地,

$$\begin{aligned}
&\mathbb{E}_{p(t,x)}[\operatorname{div}\left(\tilde{v}(x,t)-v(t,X_t)\right)] \leq \varepsilon+(t-t_k)\Big(L_t^{\operatorname{div},x}(L^{\hat{v}}+2L^{v})+L_t^{\operatorname{div},t}+L_R(L^{\hat{v}})^2 d\Big)\\
\leq&\varepsilon+(t-t_k)\Big(\frac{128\pi(d-1)^2}{(1-t)^3}\frac{M_1}{m_1}(3\pi+\varepsilon)+\frac{128\pi^2(d-1)^2}{(1-t)^3}\frac{M_1}{m_1}+(\pi+\varepsilon)^2 d\Big)\\
\lesssim&\varepsilon+(t-t_k)\Big(\frac{(d-1)^2}{(1-t)^3}\frac{M_1}{m_1}+d\Big).
\end{aligned}$$

采用提前停止时，我们在时间 $T<1$ （即$t_0=0,\ t_N=T$）终止采样算法（算法1）

$$\begin{aligned}
&\operatorname{TV}\left(\pi_T,\hat{\pi}_T\right)=\operatorname{TV}\left(X_{t_N},Y_{t_N}\right)\\
\leq&\operatorname{TV}\left(X_{t_0},Y_{t_0}\right)+\int_{t_0}^{t_N}\mathbb{E}[\|\operatorname{grad}\log p(t,x)\|\cdot\|\tilde{v}(x,t)-v(x,t)\|]dt+\int_{t_0}^{t_N}\mathbb{E}[|\operatorname{div}\left(\tilde{v}(x,t)-v(x,t)\right)|]dt\\
\lesssim&\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\varepsilon\frac{d-1}{1-t}\Big(\frac{M_1}{m_1}\Big)^{\frac{1}{2}}+(t-t_i)\frac{(d-1)^2}{(1-t)^2}\Big(\frac{M_1}{m_1}\Big)^{\frac{3}{2}}dt+\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\varepsilon+(t-t_i)\Big(\frac{(d-1)^2}{(1-t)^3}\frac{M_1}{m_1}+d\Big)dt\\
\lesssim&\Big(\frac{M_1}{m_1}\Big)^{\frac{1}{2}}(d-1)\varepsilon\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\frac{1}{1-t}dt+\Big(\frac{M_1}{m_1}\Big)^{\frac{3}{2}}(d-1)^2\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\frac{t-t_i}{(1-t)^3}dt.
\end{aligned}$$

我们首先讨论恒定步长的情形。利用

$$\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\frac{1}{1-t}dt=-\log(1-t_N)$$

以及（利用 $s:=1-t$，从而 $ds=-dt$）

$$\begin{aligned}
&\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\frac{t-t_i}{(1-t)^3}dt\leq h\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}}\frac{1}{(1-t)^3}dt=h\sum_{i=0}^{N-1}\int_{1-t_{i+1}}^{1-t_i}\frac{1}{s^3}\\
=&h\sum_{i=0}^{N-1}-\frac{1}{(1-t_i)^2}+\frac{1}{(1-t_{i+1})^2}\lesssim h\frac{1}{(1-T)^2},
\end{aligned}$$

我们最终可以将误差界定为

$$\varepsilon\Big(\frac{M_1}{m_1}\Big)^{\frac{1}{2}}(d-1)\log(\frac{1}{1-T})+h\Big(\frac{M_1}{m_1}\Big)^{\frac{3}{2}}(d-1)^2\frac{1}{(1-T)^2}.$$

现在我们有

$$\operatorname{TV}\left(X_T,Y_T\right)\lesssim\varepsilon\Big(\frac{M_1}{m_1}\Big)^{\frac{1}{2}}(d-1)\log(\frac{1}{1-T})+h\Big(\frac{M_1}{m_1}\Big)^{\frac{3}{2}}(d-1)^2\frac{1}{(1-T)^2}.$$

要获得精度为 $\varepsilon_{\text{target}}$ 的样本，我们假设 ε 足够小。于是我们需要

$$h\Big(\frac{M_1}{m_1}\Big)^{\frac{3}{2}}(d-1)^2\frac{1}{(1-T)^2}=\mathcal{O}(\varepsilon_{\text{target}}),$$

这意味着

$$h=\frac{\varepsilon_{\text{target}}(1-T)^2}{(d-1)^2}\Big(\frac{m_1}{M_1}\Big)^{\frac{3}{2}}. \tag{14}$$

接下来，我们讨论一种能够改善对 $\frac{1}{1-T}$ 依赖性的特定步长调度。记作$h_k = t_{k+1} - t_k$。那么

$$\begin{aligned}\int_{t_i}^{t_{i+1}} \frac{1-t_i}{(1-t)^3}dt =& \frac{1}{2}(1-t_i)(\frac{1}{(1-t_{i+1})^2} - \frac{1}{(1-t_i)^2}) = \frac{1}{2}(1-t_i)(\frac{t_i^2 - 2t_i - t_{i+1}^2 + 2t_{i+1}}{(1-t_{i+1})^2(1-t_i)^2}) \\ =& \frac{1}{2}(1-t_i)\frac{(t_{i+1}-t_i)(2-t_i-t_{i+1})}{(1-t_{i+1})^2(1-t_i)^2} = \frac{1}{2}(1-t_i)h_i\frac{2-t_i-t_{i+1}}{(1-t_{i+1})^2(1-t_i)^2},\end{aligned}$$

$$\int_{t_i}^{t_{i+1}} \frac{t-1}{(1-t)^3}dt = -\int_{t_i}^{t_{i+1}} \frac{1}{(1-t)^2}dt = \frac{1}{1-t_i} - \frac{1}{1-t_{i+1}} = -\frac{h_i}{(1-t_{i+1})(1-t_i)}.$$

综上，

$$\begin{aligned}\int_{t_i}^{t_{i+1}} \frac{t-t_i}{(1-t)^3}dt =& \frac{1}{2}(1-t_i)h_i\frac{2-t_i-t_{i+1}}{(1-t_{i+1})^2(1-t_i)^2} - \frac{h_i}{(1-t_{i+1})(1-t_i)} \\ =& \frac{h_i}{(1-t_{i+1})(1-t_i)}(-1 + \frac{1}{2}\frac{2-t_i-t_{i+1}}{(1-t_{i+1})}) \\ =& \frac{h_i^2}{2(1-t_{i+1})^2(1-t_i)} \lesssim \frac{h_i^2}{(1-t_{i+1})^3}.\end{aligned}$$

现在我们要控制

$$\sum_{i=0}^{N-1}\int_{t_i}^{t_{i+1}} \frac{t-t_i}{(1-t)^3}dt \lesssim \sum_{i=0}^{N-1}\frac{h_i^2}{(1-t_{i+1})^3}.$$

有 $t_i = 1 - \frac{1}{(1+\eta i)^2}$。然后 $h_i = \frac{1}{(1+\eta i)^2} - \frac{1}{(1+\eta(i+1))^2} = \frac{\eta}{(2+\eta(2i+1))(1+\eta i)^2(1+\eta(i+1))^2}$。

$$\begin{aligned}\sum_{i=0}^{N-1}\frac{h_i^2}{(1-t_{i+1})^3} =& \sum_{i=0}^{N-1}\frac{\eta^2(2+\eta(2i+1))^2}{(1+\eta i)^4(1+\eta(i+1))^4}(1+\eta(i+1))^6 \\ =& \eta^2\sum_{i=0}^{N-1}\frac{(2+\eta(2i+1))^2}{(1+\eta i)^4}(1+\eta(i+1))^2 \\ \lesssim& 4N\eta^2.\end{aligned}$$

注意，根据提前停止的构造，必有 $1 - \frac{1}{(1+\eta N)_2} = T$，这意味着$\frac{\sqrt{\frac{1}{1-T}}-1}{\eta} = N$。因此

$$\sum_{i=0}^{N-1}\frac{h_i^2}{(1-t_i)^3} \lesssim 4N\eta^2 \lesssim \eta\frac{1}{\sqrt{1-T}}.$$

我们希望达到

$$\Big(\frac{M_1}{m_1}\Big)^{\frac{3}{2}}(d-1)^2\eta\frac{1}{\sqrt{1-T}} = \mathcal{O}(\varepsilon_{\text{target}}).$$

因此我们需要

$$\eta \lesssim \Big(\frac{\varepsilon_{\text{target}}\sqrt{1-T}}{(d-1)^2}\Big).$$

因此

$$N = \frac{\frac{1}{\sqrt{1-T}} - 1}{\eta} = \mathcal{O}(\frac{\frac{1}{\sqrt{1-T}} - 1}{\frac{\varepsilon_{\text{target}}\sqrt{1-T}}{(d-1)^2}}) = \mathcal{O}(\frac{d^2}{(1-T)\varepsilon_{\text{target}}}).$$

我们注意到 ，步长调度可以通过离散时间点描述如下 :

$$t_i = 1 - \frac{1}{(1+\eta i)^2}, \qquad \text{where} \qquad \eta = \mathcal{O}\left(\frac{\varepsilon_{\text{target}}\sqrt{1-T}}{(d-1)^2}\right). \tag{15}$$

□

B.7 示例：SPD流形

我们先陈述以下结果，该结果将在附录 E 中证明。

命题 B.6。 设 M 为 对称正定矩阵（n）。假设条件 *1*。我们施加以下矩条件：存在 M_{λ_1} 使得

$$\max\left\{\mathbb{E}[d(X_1,z)^2 e^{\lambda_1 d(X_1,z)}], \mathbb{E}[e^{\lambda_1 d(X_1,z)}]\right\} \le M_{\lambda_1}, \qquad \textit{where} \qquad \lambda_1 = 24\max\{1,\kappa\}.$$

我们选择先验分布为以某个 $z \in M$ 为中心的黎曼高斯分布：p_0（x） $\propto$ 指数映射 $\big(-d\,d$（x, z）$^2\big)$。于是得到以下正则性结果。

$$\begin{aligned}
\mathbb{E}[\|v(t,x)\|^2] &\lesssim d, \\
\mathbb{E}[\|\nabla v(t,x)\|] &\lesssim \frac{d^{2+6\lambda}}{1-t} L_R M_{\lambda_1}^{\frac{1}{2}}, & \mathbb{E}[\|\nabla v(t,x)\|^2] &\lesssim \frac{d^{3+12\lambda}}{(1-t)^2} L_R^2 M_{\lambda_1}^{\frac{1}{2}}, \\
\mathbb{E}[|\frac{d}{dt} v(t,x)|] &\lesssim \frac{d^{2+6\lambda}}{1-t} L_R M_{\lambda_1}^{\frac{1}{2}}, & \mathbb{E}[|\frac{d}{dt} v(t,x)|^2] &\lesssim \frac{d^{3+12\lambda}}{(1-t)^2} L_R^2 M_{\lambda_1}^{\frac{1}{2}}, \\
\mathbb{E}[\|\operatorname{grad}_x \operatorname{div}\, v(t,x)\|] &\lesssim \frac{d^{3+12\lambda}}{(1-t)^2} L_R^3, & \mathbb{E}[\|\operatorname{grad}_x \operatorname{div}\, v(t,x)\|^2] &\lesssim \frac{d^{5+24\lambda}}{(1-t)^4} L_R^6 M_{\lambda_1}, \\
\mathbb{E}[|\frac{d}{dt} \operatorname{div}\, v(t,x)|] &\lesssim \frac{d^{3+12\lambda}}{(1-t)^2} L_R^3 M_{\lambda_1}^{\frac{1}{2}},
\end{aligned}$$

其中 $\lambda =$ 最大值$\{1,\kappa\}$。

证明。 [命题 4 的证明1] 利用命题 B.6，并根据定理 2，我们有

$$\begin{aligned}
&\operatorname{TV}(\pi_T, \hat{\pi}_T) = \operatorname{TV}(X_{t_N}, Y_{t_N}) \\
\le\ &\operatorname{TV}(X_{t_0}, Y_{t_0}) + \frac{1}{2} N h^2 \Big(\sqrt{6 L_t^{\text{score}}} \left((L_t^{\hat{v},x})^2 (L_t^{\hat{v}})^2 + (L_t^{\hat{v},x})^2 (L_t^v)^2 + (L_t^{v,t})^2 + (L_t^v)^2 (L_t^{v,x})^2 \right) \\
&+ \Big(L_t^{\operatorname{div}\,\hat{v},x} (L_t^{\hat{v}} + L_t^v) + L_t^{\operatorname{div},x} L_t^v + L_t^{\operatorname{div},t} + L_R (L_t^{\hat{v}})^2 d \Big) \Big) + N h (\varepsilon \sqrt{2 L_t^{\text{score}}} + \varepsilon) \\
\lesssim\ & N h^2 \Big(\frac{d^{5+12\lambda}}{(1-t)^2} L_R^2 M_{\lambda_1} \sqrt{M \frac{d^2}{(1-t)^2} L_R^2 d^{12\lambda}} \Big) + N h (\varepsilon \sqrt{M \frac{d^2}{(1-t)^2} L_R^2 d^{12\lambda}} + \varepsilon) \\
\lesssim\ & h \Big(\frac{d^{6+18\lambda}}{(1-t)^3} L_R^3 M_{\lambda_1}^{\frac{3}{2}} \Big) + (\frac{d^{6\lambda+1}}{1-t} L_R) M_{\lambda_1}^{\frac{1}{2}} \varepsilon.
\end{aligned}$$

因此，要达到 $\varepsilon_{\text{target}}$ 精度，我们需要

$$h = \mathcal{O}\left(\frac{\varepsilon_{\text{target}}(1-T)^3}{d^{6+18\lambda}L_R^3 M_{\lambda_1}^{\frac{3}{2}}}\right).$$

因此，迭代复杂度为

$$N = \mathcal{O}\left(\frac{d^{6+18\lambda}L_R^3 M_{\lambda_1}^{\frac{3}{2}}}{\varepsilon_{\text{target}}(1-T)^3}\right).$$

我们注意到，由于 $K_{\min} = -\frac{1}{2}$（Criscitiello 和 Boumal，2023年），因此 $\lambda = \max\{1, \kappa\} = 1$。与此同时，引理 5.2 所要求的 h 上界可以降低至

$$\min\{\frac{1}{L_\nabla}, \sqrt{\frac{1}{\|b(x)\|^2 L_R L_\nabla}}\} = \frac{1-T}{d^8 L_R M_{\lambda_1}^{\frac{1}{2}}},$$

这低于我们所要求的 h阶。

因此，我们得到迭代复杂度为

$$N = \mathcal{O}\left(\frac{d^{6+18\lambda}L_R^3 M_{\lambda_1}^{\frac{3}{2}}}{\varepsilon_{\text{target}}(1-T)^3}\right) = \mathcal{O}\left(\frac{d^{24}L_R^3 M_{\lambda_1}^{\frac{3}{2}}}{\varepsilon_{\text{target}}(1-T)^3}\right).$$

□

C 主要定理证明的辅助结果

C.1 雅可比方程

引理 C.1 (雅可比方程的解)。 设 M 具有常数截面曲率 c。设 J 是沿着 γ的法向雅可比场，其初始条件为 $J(0) = v^\perp, J'(0) = 0$。则有

$$J(t) = s_c^{(2)}(t)\|v^\perp\|E(t) := \begin{cases} \|v^\perp\|E(t), & \text{if } c = 0, \\ \|v^\perp\|\cos(\sqrt{c}t)E(t), & \text{if } c > 0, \\ \|v^\perp\|\cosh(\sqrt{-c}t)E(t), & \text{if } c < 0. \end{cases}$$

其中， E 是平行法向单位向量场，且 $E(0) = \frac{v^\perp}{\|v^\perp\|}$。 ——

证明。 [引理 C. 的证明1] 与 Lee（2018，命题 10.12）的证明类似，解的形式为 $J(t) = f(t)E(t)$, 其中 $E(t)$ 是沿着 γ的平行法向单位向量场。由于曲率为常数，且 J 是法向雅可比场，雅可比方程可化简为 $D_t^2 J + cJ = 0$，因此我们只需对 $f''(t) + cf(t) = 0$ 求解，其中 $f(t) \in \mathbb{R}, \forall t$。

- 当 $c = 0$ 时，我们得到 $f''(t) = 0$，因此（为满足初始条件） $f(t) = \|v^\perp\|$。
- 当 $c > 0$ 时，我们得到 $f(t) = \|v^\perp\|\cos(\sqrt{c}t)$。
- 当 $c < 0$ 时，我们得到 $f(t) = \|v^\perp\|\frac{e^{\sqrt{-c}t}+e^{-\sqrt{-c}t}}{2} = \|v^\perp\|\cosh(\sqrt{c}t)$。

注意，$E(0) = \frac{v^\perp}{\|v^\perp\|}$。因此，

$$J(t) = \begin{cases} \|v^\perp\| E(t), & \text{if } c = 0, \\ \|v^\perp\| \cos(\sqrt{c}t) E(t), & \text{if } c > 0, \\ \|v^\perp\| \cosh(\sqrt{-c}t) E(t), & \text{if } c < 0. \end{cases}$$

□

引理 C.2. 设 $J_v(t)$ 是沿着 $\gamma(t) =$ 指数映射 $x(thb(x))$ 的雅可比场，其中 $J_v(0) = v$, $J_v'(0) = \omega$。定义 $Y(t) = P_{\gamma\ t}^{\gamma(0)}() J_v(t)$。在 x 的第一个共轭点之前，我们有

$$\|Y(t)\| \le \|v^\parallel\| + \|\omega^\parallel\| t + s_{K_{\min}}(t)\|\omega^\perp\| + \frac{s_{K_{\min}}(t)}{s_{K_{\min}}(t_0)}\|v^\perp\|.$$

取 $h < \frac{R}{\|b\ x'()\|}$，我们有

$$\int_0^1 \|Y(t)\| dt \le \begin{cases} \|v^\parallel\| + \|\omega^\parallel\| + \frac{\|\omega^\perp\|}{\sqrt{K_{\min}}} + \|v^\perp\|, & \text{if } K_{\min} > 0, \\ \|v^\parallel\| + \|\omega^\parallel\| + \frac{\cosh(\sqrt{-K_{\min}})-1}{-K_{\min}}\|\omega^\perp\| + \frac{\sinh(\sqrt{-K_{\min}})}{\sqrt{-K_{\min}}}\|v^\perp\|, & \text{if } K_{\min} < 0, \\ \|v^\parallel\| + \frac{1}{2}\|\omega^\parallel\| + \frac{1}{2}\|\omega^\perp\| + \|v^\perp\|, & \text{if } K_{\min} = 0. \end{cases}$$

证明。 [引理的证明 C.2] 利用平行移动的等距性质，$\|Y(t)\| = \|J_v(t)\|$。将 $v = v^\parallel + v^\perp$ 分解为 $v^\parallel$ 沿 $\gamma(0)$ 方向的分量，且 $\gamma'(0)\ \langle v^\perp,\ \gamma'(0)\rangle = 0$。类似地，我们将 $\omega = \omega^\parallel + \omega^\perp$ 分解。

定义 $J_v =: J^{(0)} + J_{(1)} + J_{(2)}$，其中

$$\begin{aligned} J^{(0)}(0) &= v^\parallel, \quad D_t J^{(0)}(0) = \omega^\parallel; \\ J^{(1)}(0) &= 0, \quad D_t J^{(1)}(0) = \omega^\perp; \\ J^{(2)}(0) &= v^\perp, \quad D_t J^{(2)}(0) = 0. \end{aligned}$$

注意到 $J_{(1)}(0)$、$D_t J_{(1)}(0)$ 都与 $\gamma'(0)$ 正交，因此 $J^{(1)}(t)$ 是一个法向雅可比场。

我们知道 $J^{(0)}$ 是一个切向雅可比场，因此它具有如下形式

$$J^{(0)}(t) = (a + bt)\gamma'(t).$$

代入初始值，得到（注意 $D_t\gamma'(0) = 0$）

$$a\gamma'(0) = v^\parallel, \quad b\gamma'(0) = \omega^\parallel.$$

因此 $|a| = \frac{\|v^\parallel\|}{\|\gamma'\ 0\ ()\|}$，且 $|b| = \frac{\|\omega^\parallel\|}{\|\gamma'\ 0\ \|()}$。

根据定义，$\|\gamma'(0)\| = \|hb(x)\| = \|\gamma'(t)\|$。于是得到

$$\|J^{(0)}(t)\| \le \|v^\parallel\| + \|\omega^\parallel\| t.$$

若 M 的所有截面曲率都不小于常数 $K_{\min}$，则雅可比场比较定理给出

$$\|J^{(1)}(t)\| \le s_K(t)\|D_t J^{(1)}(0)\| = s_{K_{\min}}(t)\|\omega^\perp\|.$$

现在考虑这一项 $J^2(0)$。由 Cheeger等人 (1975，定理1.34），并应用引理 C.1中给出的雅可比场公式，可得 $\|J^{(2)}(t)\| \leq \|\tilde{J}(t)\|$。我们指出，无焦点条件意味着 $J^{(2)}(t) \neq 0$。只要尚未到达割迹，测地线就是极小的。因此，只需确保测地线 γ 满足 $\|\gamma'(0)\| < \text{inj}\,(M)$。

最后，回忆一下 $\|Y(t)\| = \|J_v(t)\|$。

$$\begin{aligned}\|Y(t)\| =&\|J_v(t)\| = \|J^{(0)}(t) + J^{(1)}(t) + J^{(2)}(t)\| \\ \leq&\|v^{\|}\| + \|\omega^{\|}\|t + |s_{K_{\min}}(t)|\|\omega^{\perp}\| + |s^{(2)}_{K_{\min}}(t)|\|v^{\perp}\|.\end{aligned}$$

我们分 成几种情况。第一种情况是 $K_{\min} > 0$。在这种情况下，我们将 $R = \text{inj}\,(M)$ 记为..... d

$$\int_0^1 |s_{K_{\min}}(t)|dt = \frac{1}{\sqrt{K_{\min}}}\int_0^1 |\sin(t\sqrt{K_{\min}})|dt \leq \frac{1}{\sqrt{K_{\min}}}.$$

此外，

$$\int_0^1 |s^{(2)}_{K_{\min}}(t)|dt = \int_0^1 |\cos(\sqrt{K_{\min}}t)|dt \leq 1.$$

总之，当 $h < {}^{R}_{\|b\,x}\,(\,)\,\|$ 时,——

$$\int_0^1 \|Y(t)\|dt \leq \|v^{\|}\| + \|\omega^{\|}\| + \frac{\|\omega^{\perp}\|}{\sqrt{K_{\min}}} + \|v^{\perp}\|.$$

第二种情况是 $K_{\min} < 0$。
于是 $s_{K_{\min}}(t) = \frac{1}{\sqrt{-K_{\min}}}\sinh(\sqrt{-K_{\min}}t)$。考虑 $0 < t \leq h < 1$。

$$\int_0^1 s_{K_{\min}}(t)dt = \int_0^1 \frac{1}{\sqrt{-K_{\min}}}\sinh t\sqrt{-K_{\min}}dt = \frac{1}{-K_{\min}}(\cosh(\sqrt{-K_{\min}}) - 1).$$

对于第二个积分，

$$\int_0^1 \cosh(\sqrt{-K_{\min}}t)dt = \frac{\sinh(\sqrt{-K_{\min}})}{\sqrt{-K_{\min}}}.$$

综上，当 $h < {}^{R}_{\|b\,x}\,(\,)\,\|$，——

$$\int_0^1 \|Y(t)\|dt \leq \|v^{\|}\| + \|\omega^{\|}\| + \frac{\cosh(\sqrt{-K_{\min}}) - 1}{-K_{\min}}\|\omega^{\perp}\| + \|v^{\perp}\|\frac{\sinh(\sqrt{-K_{\min}})}{\sqrt{-K_{\min}}}.$$

第三种情况是 $K_{\min} = 0$。在这种情况下， $s_{K_{\min}}(t) = t$。

$$\begin{aligned}\|Y(t)\| =&\|J_v(t)\| = \|J^{(0)}(t) + J^{(1)}(t) + J^{(2)}(t)\| \\ \leq&\|v^{\|}\| + \|\omega^{\|}\|t + \|\omega^{\perp}\|t + \|v^{\perp}\|.\end{aligned}$$

于是我们有

$$\int_0^1 \|Y(t)\|dt \leq \|v^{\|}\| + \frac{1}{2}\|\omega^{\|}\| + \frac{1}{2}\|\omega^{\perp}\| + \|v^{\perp}\|.$$

□

C.2 散度项

引理 *C.3.* 令 c 表示满足 $c(0) = z$, $c(h) = x$的测地线。定义测地线变分$\Lambda(s,t)$为 $\Lambda(s,t) =$指数映射 $c(t)(sP_{c\ 0}^{c(t)}(\ )E_i(t))$，其中 $E_i(t)$ 是沿 c的基向量场。令 V为一个向量场。对于每个 z，我们都可以通过平行移动得到一个新的向量场，记为 $P_{c\ 0}^{c(t)}(\ )V(z)$。于是

$$D_s P_{c(0)}^{c(h)} V(z) = P_{c(0)}^{c(h)} D_s V(z) - P_{c(0)}^{c(h)} \int_0^h P_{c(\tau)}^{c(t_k)} R(\partial_s \Lambda(0,\tau), \partial_t \Lambda(0,\tau)) P_{c(t_k)}^{c(\tau)} V(z) d\tau,$$

其中 D_s 表示沿 $E_i(t)$的协变导数。

注意， s 方向实际上是任意的。我们的目标是计算某一点的散度，因此可以遍历所有可能的 s 方向，即所有基向量。粗略来说，每个 s 方向都定义了一条 D_s。

证明。 [引理 C3] 根据构造，有

$$\Lambda(0,0) = z, \Lambda(0,h) = x.$$

且平行移动 $P_z^x = P_\Lambda^{\Lambda(0,h)}{}_{(0,0)}$沿着 c进行。定义 $W(s,t) = P_{\Lambda\ s\ 0}^{\Lambda(s,t)}{}_{(,)} V(\Lambda(s,0))$，其中我们沿曲线 $t \mapsto \Lambda(s,t)$进行平行移动，该曲线可能不是测地线。根据平行移动的定义， $D_t W(s,t) = 0, \forall s$。

我们知道

$$-D_t D_s W(s,t) = D_s D_t W(s,t) - D_t D_s W(s,t) = R(\partial_s \Lambda(s,t), \partial_t \Lambda(s,t)) W(s,t).$$

因此，在 $s = 0$ 处求值可得

$$D_t D_s W(0,t) = -R(\partial_s \Lambda(0,t), \partial_t \Lambda(0,t)) W(0,t).$$

我们对等式两边进行平行移动 $P_{c\ t}^{c(0)}{}_{()}$，根据引理 C.5 可得

$$\frac{d}{dt} P_{c(t)}^{c(0)} D_s W(0,t) = P_{c(t)}^{c(0)} D_t D_s W(0,t) = -P_{c(t)}^{c(0)} R(\partial_s \Lambda(0,t), \partial_t \Lambda(0,t)) W(0,t).$$

注意，等式两边都是 $T_{c(0)}M = T_z M$ 上的时变向量场。因此，我们可以进行积分

$$\int_0^h \frac{d}{d\tau} P_{c(\tau)}^{c(t_k)} D_s W(0,\tau) d\tau = -\int_0^h P_{c(\tau)}^{c(t_k)} R(\partial_s \Lambda(0,\tau), \partial_t \Lambda(0,\tau)) W(0,\tau) d\tau.$$

因此

$$P_{c(h)}^{c(0)} D_s W(0,h) = D_s W(0,0) - \int_0^h P_{c(\tau)}^{c(t_k)} R(\partial_s \Lambda(0,\tau), \partial_t \Lambda(0,\tau)) W(0,\tau) d\tau.$$

注意 $D_s W(0,\ 0) = D_s V(\Lambda(0,\ 0))$。对等式两边进行平行移动 $P_{c\ 0}^{c(h)}{}_{()}$，得到

$$D_s W(0,h) = P_{c(0)}^{c(h)} D_s V(\Lambda(0,0)) - P_{c(0)}^{c(h)} \int_0^h P_{c(\tau)}^{c(t_k)} R(\partial_s \Lambda(0,\tau), \partial_t \Lambda(0,\tau)) W(0,\tau) d\tau.$$

回忆一下 $W(s,\ t) = P_{\Lambda\ s\ 0}^{\Lambda(s,t)}(\cdot\ )V(\Lambda(s,\ 0))$。

$$D_s P_{c(0)}^{c(h)} V(\Lambda(0,0)) = P_{c(0)}^{c(h)} D_s V(\Lambda(0,0)) - P_{c(0)}^{c(h)} \int_0^h P_{c(\tau)}^{c(t_k)} R(\partial_s \Lambda(0,\tau), \partial_t \Lambda(0,\tau)) P_{c(t_k)}^{c(\tau)} V(z) d\tau.$$

□

注意，$\tilde{v}$是通过$\hat{v}$和 F的逆定义的。在分析中，我们需要控制 div $\tilde{v}(x,\ t)$，方法是将其写成涉及 div $\hat{v}$的某种表达式。

引理 C.4。 在假设 *1*下，我们有

$$|\operatorname{div}(\tilde{v}(x,t) - v(x,t))| \le |\operatorname{div}\ \hat{v}(t_k,z) - \operatorname{div}\ v(x,t)| + L_R(t-t_k) d \|\hat{v}(t_k,z)\|^2.$$

其中 指数映射$_z((t-t_k)\hat{v}(t_k,\ z))\ = x$。

证明。 [引理 C.4] 我们沿用引理 C.3 中的设定，其中 V 被 $\hat{v}$替换。我们用$\{E_i(t)\}_{i=1}^d$ 表示沿测地线 c的一个正交归一基向量场。请注意，一个细微的区别是时间平移：在引理 C.3 中，曲线的时间范围是从 0 到 h，而这里是从 t_k到 t （因此有 $h = t - t_k$）。我们将时间变量记为 τ。在大多数情况下，我们指的是变量τ 属于 $[t_k, t]$。

根据定义，指数映射 $_z((t-t_k)\hat{v}(t_k,\ z))\ = x$。因此我们知道 $\Lambda(0,\ \tau)$ =指数映射 $_z((\tau - t_k)\hat{v}(t_k,\ z))$。根据构造，

$$\partial_t \Lambda(0,\tau) = P_{c(t_k)}^{c(\tau)} \hat{v}(t_k,z), \quad \partial_s \Lambda(0,\tau) = E_i(\tau).$$

根据引理 C.3，结合 $c\ (\tau)$ =指数映射 $_z\ (\ (\tau - t_k)\ \hat{v}\ (t_k,\ z)\ )$，

$$D_s P_{c(t_k)}^{c(t)} \hat{v}(t_k,z) = P_{c(t_k)}^{c(t)} D_s \hat{v}(t_k,z) - P_{c(t_k)}^{c(t)} \int_{t_k}^t P_{c(\tau)}^{c(t_k)} R(E_i(\tau), P_{c(t_k)}^{c(\tau)} \hat{v}(t_k,z)) P_{c(t_k)}^{c(\tau)} \hat{v}(t_k,z) d\tau.$$

根据黎曼散度的定义，

$$\begin{aligned}
\operatorname{div}\ \tilde{v}(t,x) &= \sum_{i=1}^d \langle \nabla_{E_i(t)} \tilde{v}(t,x), E_i(t)\rangle = \sum_{i=1}^d \langle \nabla_{E_i(t)} P^x_{F^{-1}_{t_k,t-t_k}(x)} \hat{v}(t_k, F^{-1}_{t_k,t-t_k}(x)), E_i(t)\rangle \\
&= \sum_{i=1}^d \langle D_s^{(i)} P_{c(t_k)}^{c(t)} \hat{v}(t_k,z), E_i(t)\rangle \\
&= \sum_{i=1}^d \langle P_{c(t_k)}^{c(t)} D_s^{(i)} \hat{v}(t_k,z) - P_{c(t_k)}^{c(t)} \int_{t_k}^t P_{c(\tau)}^{c(t_k)} R(E_i(\tau), P_{c(t_k)}^{c(\tau)} \hat{v}(t_k,z)) P_{c(t_k)}^{c(\tau)} \hat{v}(t_k,z) d\tau, E_i(t)\rangle \\
&= \sum_{i=1}^d \langle P_{c(t_k)}^{c(t)} \nabla_{E_i(t_k)} \hat{v}(t_k,z) - P_{c(t_k)}^{c(t)} \int_{t_k}^t P_{c(\tau)}^{c(t_k)} R(E_i(\tau), P_{c(t_k)}^{c(\tau)} \hat{v}(t_k,z)) P_{c(t_k)}^{c(\tau)} \hat{v}(t_k,z) d\tau, E_i(t)\rangle \\
&= \sum_{i=1}^d \langle P_{c(t_k)}^{c(t)} \nabla_{E_i(t_k)} \hat{v}(t_k,z), E_i(t)\rangle \\
&\quad - \sum_{i=1}^d \langle P_{c(t_k)}^{c(t)} \int_{t_k}^t P_{c(\tau)}^{c(t_k)} R(E_i(\tau), P_{c(t_k)}^{c(\tau)} \hat{v}(t_k,z)) P_{c(t_k)}^{c(\tau)} \hat{v}(t_k,z) d\tau, E_i(t)\rangle
\end{aligned}$$

$$= \text{div } \hat{v}(t_k, z) - \sum_{i=1}^{d} \langle P_{c(t_k)}^{c(t)} \int_{t_k}^{t} P_{c(\tau)}^{c(t_k)} R(E_i(\tau), P_{c(t_k)}^{c(\tau)} \hat{v}(t_k, z)) P_{c(t_k)}^{c(\tau)} \hat{v}(t_k, z) d\tau, E_i(t) \rangle,$$

其中 $D_s^{(i)}$表示对应于 $E_i(\tau)$ 的协变导数。我们有

$$\begin{aligned}
&\sum_{i=1}^{d} \langle P_{c(t_k)}^{c(t)} \int_{t_k}^{t} P_{c(\tau)}^{c(t_k)} R(E_i(\tau), P_{c(t_k)}^{c(\tau)} \hat{v}(t_k, z)) P_{c(t_k)}^{c(\tau)} \hat{v}(t_k, z) d\tau, E_i(t) \rangle \\
=&\sum_{i=1}^{d} \int_{t_k}^{t} \langle P_{c(\tau)}^{c(t_k)} R(E_i(\tau), P_{c(t_k)}^{c(\tau)} \hat{v}(t_k, z)) P_{c(t_k)}^{c(\tau)} \hat{v}(t_k, z), E_i(t_k) \rangle d\tau \\
=&\sum_{i=1}^{d} \int_{t_k}^{t} \langle R(E_i(\tau), P_{c(t_k)}^{c(\tau)} \hat{v}(t_k, z)) P_{c(t_k)}^{c(\tau)} \hat{v}(t_k, z), E_i(\tau) \rangle d\tau \\
\leq&\sum_{i=1}^{d} \int_{t_k}^{t} L_R \|\hat{v}(t_k, z)\|^2 d\tau \leq (t - t_k) d L_R \|\hat{v}(t_k, z)\|^2.
\end{aligned}$$

由此可得

$$\begin{aligned}
&|\text{div } (\tilde{v}(x,t) - v(x,t))| \\
=&|\text{div } \hat{v}(t_k, z) - \sum_{i=1}^{d} \langle P_{c(t_k)}^{c(t)} \int_{t_k}^{t} P_{c(\tau)}^{c(t_k)} R(E_i(\tau), P_{c(t_k)}^{c(\tau)} \hat{v}(t_k, z)) P_{c(t_k)}^{c(\tau)} \hat{v}(t_k, z) d\tau, E_i(t) \rangle - \text{div } v(x,t)| \\
\leq&|\text{div } \hat{v}(t_k, z) - \text{div } v(x,t)| + L_R (t - t_k) d \|\hat{v}(t_k, z)\|^2.
\end{aligned}$$

□

引理 C.5。我们有

$$\frac{d}{dt} P_{c(t)}^{c(0)} Y(t) = P_{c(t)}^{c(0)} D_t Y(t).$$

证明。 [引理 C.5] 将沿 t 方向的测地线记为 $c(t)$。任取 $v \in T_{c(0)} M$，并定义 $Z(t) = P_{c\ 0}^{c(t)}{}_{()} v$。

$$\frac{d}{dt} \langle P_{c(t)}^{c(0)} Y(t), v \rangle = \frac{d}{dt} \langle Y(t), P_{c(0)}^{c(t)} v \rangle = \langle D_t Y(t), P_{c(0)}^{c(t)} v \rangle + 0 = \langle D_t Y(t), P_{c(0)}^{c(t)} v \rangle = \langle P_{c(t)}^{c(0)} D_t Y(t), v \rangle.$$

另一方面，

$$\frac{d}{dt} \langle P_{c(t)}^{c(0)} Y(t), v \rangle = \langle \frac{d}{dt} P_{c(t)}^{c(0)} Y(t), v \rangle.$$

由于上述结论对任意 v都成立，因此我们有

$$\frac{d}{dt} P_{c(t)}^{c(0)} Y(t) = P_{c(t)}^{c(0)} D_t Y(t).$$

□

D 超球面正则性结果

回忆一下，

$$v(t,x)=\frac{1}{1-t}\int_M \mathrm{Log}_x(x_1)\,p_t(x_1|x)\,dV_g(x_1).$$

为了建立 v 的正则性，自然需要研究条件密度公式 $p_t(x_1|x)$。在欧几里得空间中，以高斯分布作为先验分布 p_0时，我们有 $p_t(x_1 \mid x_t = x) \propto p_1(x_1)\exp(-\frac{\|tx_1-x\|^2}{2\,1-t^{\,2}}())$。这是流匹配中的一个标准结果，例如参见 Guan 等人(2026年) 和 Zhou 和 Liu (2025年)。然而，在黎曼流形上，曲率会因变量替换公式而引入一个额外项，而割点的存在则会引入一个额外的示性函数。我们给出以均匀分布为先验分布、采用测地线插值时超球面 $\mathcal{S}^d$上的条件密度公式。见下文的引理 D.1。

引理 D.1 (S^d 上测地线插值的条件密度)。令 S^d 为配备圆度量的单位球面， $d \geq 2$。令 $X_1 \sim p_1$ 为具有光滑密度 $p_1 > 0$ （相对于 dV_g）的数据分布，且 $X_0 \sim p_0$ 为与 X_1相互独立的均匀分布。考虑采用最短测地线的测地线插值。固定 $t \in [0,1)$ 以及 $x, x_1 \in S^d$。记 $r = d(x, x_1)$。记

$$\begin{aligned} J_t(x \mid x_1) &= \frac{1}{1-t}\Big(\frac{\sin\big(r/(1-t)\big)}{\sin r}\Big)^{d-1}\mathbf{1}_{\{d(x,x_1)<(1-t)\pi\}} \\ &= \begin{cases} \frac{1}{1-t}\Big(\frac{\sin\big(r/(1-t)\big)}{\sin r}\Big)^{d-1}, & \text{if } r<(1-t)\pi, \\ 0, & \text{if } r\geq(1-t)\pi. \end{cases} \end{aligned}$$

那么， X_1 在给定 $X_t = x$ 条件下的条件密度为

$$p_t(x_1 \mid x)=\frac{p_1(x_1)\,J_t(x \mid x_1)}{\displaystyle\int_{S^d} p_1(z)\,J_t(x \mid z)dV_g(z)}.$$

证明。 证明思路如下。我们首先写出 X_0, X_1的联合分布，然后使用变量替换公式得到 X_t, X_1的联合分布。

我们应用变量替换公式。我们需要在(X_0, X_1)与(X_t, X_1)之间建立一个微分同胚。固定 $t \in (0,1)$和 $x_1 \in S^d$。在 x_1处使用极坐标：每个 $y \in S^d \setminus \{-x_1\}$ 都可以唯一地写成 $y =$指数映射$_{x_1}(r\omega)$，其中 $r \in (0,\pi)$ $\omega \in S^{d-1}$。我们知道黎曼体积元为

$$dV_g(y) = (\sin r)^{d-1} dr\, d\omega.$$

特别地，对 于任意 x_0，我们令 $r_0 = d(x_0, x_1) \in (0,\pi)$，并令 $\omega =$对数映射$_{x_1(x_0)}\|$对数映射$_{x_1(x_0)}\|_{\circ}$ 于是我们可以写 e

$$x_0 = \mathrm{Exp}_{x_1}(r_0\omega), \qquad x_1 = \mathrm{Exp}_{x_1}(0\cdot\omega).$$

and

$$X_t = \mathrm{Exp}_{x_0}\big(t\,\mathrm{Log}_{x_0}(x_1)\big) = \mathrm{Exp}_{x_1}\big((1-t)r_0\,\omega\big).$$

T因此，我们可以将所需的微分同胚定义为 F_t： $(r_0,\ \omega) \mapsto (r,\ \omega)$，满足

$$F_{t,x_1}(r_0,\omega) = ((1-t)r_0,\omega), \qquad F^{-1}_{t,x_1}(r,\omega) = (\frac{r}{1-t},\omega).$$

但请注意，我们必须限制 $r < (1-t)\pi$，否则 $\frac{r}{1-t} \notin (0, \pi)$，因此$F_{t,x_1}^{-1}(r, \omega)$ 将不再处于极坐标下。

现在回顾变量替换公式。我们应有

$$\int_{S^d} p_0(F_{t,x_1}^{-1}(x))|\det dF_{t,x_1}^{-1}|dV_g(x) = \int_{S^d} p_0(x_0)dV_g(x_0).$$

用极坐标表示：x_0 和 x 处的体积元为

$$dV_g(x_0) = (\sin r_0)^{d-1}dr_0\, d\omega, \qquad dV_g(x) = (\sin r)^{d-1}dr\, d\omega.$$

定义函数 $J_t(x \mid x_1)$，使其满足

$$dV_g(x_0) = J_t(x \mid x_1)\, dV_g(x).$$

利用 $r_0 = r/(1-t)$（因此有 $dr_0 = dr/(1-t)$），可得

$$(\sin r_0)^{d-1}dr_0\, d\omega = J_t(x \mid x_1)(\sin r)^{d-1}dr\, d\omega$$
$$(\sin(r/(1-t)))^{d-1}\frac{dr}{1-t}d\omega = J_t(x \mid x_1)(\sin r)^{d-1}dr\, d\omega,$$

现在我们推导密度。注意，根据我们的测地线插值构造，对于任意x_0 和 t，所得的 x 必须满足 $d(x_1, x) = (1-t)d(x_1, x_0) \le (1-t)\pi$。因此

$$J_t(x \mid x_1) = \frac{1}{1-t}\Big(\frac{\sin\big(r/(1-t)\big)}{\sin r}\Big)^{d-1}, \qquad r = d(x, x_1),$$

对所有 $r < (1-t)\pi$，且在其他情况下 $J_t(x \mid x_1) = 0$。因此，我们可以等价地写成

$$J_t(x \mid x_1) = \frac{1}{1-t}\Big(\frac{\sin\big(r/(1-t)\big)}{\sin r}\Big)^{d-1}\mathbf{1}_{\{d(x,x_1)<(1-t)\pi\}}.$$

(X_0, X_1) 的联合密度为

$$p(x_0, x_1) = p_0(x_0)p_1(x_1)$$

关于 $dV_g(x_0)\, dV_g(x_1)$ 的密度。根据变量替换公式，

$$p_t(x, x_1) = p_0(F_{t,x_1}^{-1}(x))p_1(x_1)J_t(x \mid x_1).$$

最后，由于 $p_t(x, x_1) = p_t(x_1 \mid x)p_t(x) = p_t(x_1 \mid x)\int_{S^d} p_t(x, z)\, dV_g(z)$，$X_1$ 给定 $X_t = x$ 的条件密度为

$$p_t(x_1 \mid x) = \frac{p_t(x, x_1)}{\int_{S^d} p_t(x, z)\, dV_g(z)},$$

由此得到引理中所述的公式。

□

根据引理 D.1，我们可以推导出插值密度的以下公式。

引理 D.2. 我们可以将插值密度写作

$$p_t(x) = \int_{S^d} p_t(x, x_1)\, dV_g(x_1) = \frac{1}{\mathrm{Vol}\,(S^d)}\int_{S^d} p_1(x_1)J_t(x \mid x_1)dV_g(x_1).$$

证明. 根据引理D.1，我们有

$$p_t(x, x_1) = p_0(F_{t,x_1}^{-1}(x))p_1(x_1)J_t(x \mid x_1).$$

因此，我们得到（注意， p_0 服从均匀分布）

$$p_t(x) = \int_{S^d} p_t(x, x_1)\, dV_g(x_1) = \frac{1}{\mathrm{Vol}\,(S^d)} \int_{S^d} p_1(x_1)J_t(x \mid x_1)dV_g(x_1).$$

□

在论证 v 的正则性之前，我们首先需要说明示性函数（对应于割点）不会破坏其光滑性。在下面的引理中，我们证明，在超球面上，正弦函数（出现在条件密度函数中）会使示性函数变得光滑，从而得到一个光滑向量场。

引理 D.3. 设 $\mathcal{S}^d$ 为配备圆度量 $d \geq 2$ 的单位超球面。设 $x_1 \in \mathcal{S}^d$，并定义，对于$t \in [0,1)$和 $x \in \mathcal{S}^d$。将 $r(x) = d(x, x_1)$ 记为径向距离函数。那么，对于满足 $0 \leq m \leq d-2$ 的整数 m，函数 $J(t, x \mid x_1)$ 视为定义在 $[0,1) \times \mathcal{S}^d$ 上的函数时，是 C^m。

因此，作为条件期望的 $v(t,x)$ 在(t,x) 中关于 $t \in [0,1)$ 是 C^2。由此，流匹配常微分方程的解存在且唯一。

证明。 记 $\mathcal{U} := \{(t, x) \in [0,1) \times \mathcal{S}^d : r(x) < (1-t)\pi\}$。我们首先证明其在$\mathcal{U}$上的光滑性。固定 $(t_0, x_0) \in \mathcal{U}$，其中 $x_0 \neq x_1$，并记 $r_0 := r(x_0) > 0$。由于 $r_0 < (1-t_0)\pi < \pi$，我们有$x_0 \neq -x_1$。此外，由于 $r(\cdot)$ 的连续性，存在 (t_0, x_0) 的一个邻域 $\mathcal{V}$，使得对所有 $(t, x) \in \mathcal{V}$， $\mathbf{1}_{\{r(x)<(1-t)\pi\}} \equiv 1$，因此

$$J(t, x \mid x_1) = \frac{1}{1-t}\Big(\frac{\sin\big(r(x)/(1-t)\big)}{\sin r(x)}\Big)^{d-1}.$$

由于当 $x \notin \{x_1, -x_1\}$ 时， $r = d(x, x_1)$ 是光滑的， J作为光滑函数的复合函数也是光滑的。接下来需要检查 $J(t, x \mid x_1)$ 在点 $x = x_1 (r = 0)$ 处的光滑性。在 x_1 处引入正规坐标：对于 x_1 附近的 x，写作

$$v := \mathrm{Log}_{x_1}(x) \in T_{x_1}\mathcal{S}^d, \qquad r(x) = d(x, x_1) = \|v\|.$$

只需证明在正规坐标下，作为 v 的函数的 J 是光滑的。注意到 $\|v\|$ 在 $v = 0$ 处不可微，因此像 $v \mapsto \sin\|v\|$ 这样的函数在 0 处并不 C^1。但请注意，在 J 中， $\|v\|$ 上的“非光滑”部分会相互抵消。通过检查光滑性

$$\frac{\sin(\|v\|/(1-t))}{\sin\|v\|},$$

我们得出 J 在 U 上的光滑性。

我们通过检查边界 $r = (1-t)\pi$ 来推广光滑性结果。为证明 C^m，固定 (t_0, x_0)，其中 $r(x_0) = (1-t_0)\pi$。由于 $r(x_0) \in (0, \pi)$，我们知道 $x \mapsto r(x)$ 在 x_0 附近光滑。将 $s(t, x)$ 定义如下，以描述其到达边界的距离：

$$s(t, x) := (1-t)\pi - r(x).$$

在侧面 $s > 0$，我们有 $u := r/(1-t) \in (0, \pi)$ 以及

$$\sin\Big(\frac{r}{1-t}\Big) = \sin\Big(\pi - \frac{s}{1-t}\Big) = \sin\Big(\frac{s}{1-t}\Big).$$

因此，对于 $s > 0$,

$$J(t,x \mid x_1) = \frac{1}{1-t}\Big(\frac{\sin\big(s(t,x)/(1-t)\big)}{\sin r(x)}\Big)^{d-1} = s(t,x)^{d-1}(1-t)^{-d}\left(\frac{\frac{\sin\left(s(t,x)/(1-t)\right)}{s(t,x)/(1-t)}}{\sin r(x)}\right)^{d-1}.$$

在整个流形 $\mathcal{S}^d$ 上，我们有

$$J(t,x \mid x_1) = \big(s(t,x)\big)_+^{d-1}(1-t)^{-d}\left(\frac{\frac{\sin\left(s(t,x)/(1-t)\right)}{s(t,x)/(1-t)}}{\sin r(x)}\right)^{d-1}, \qquad (s)_+ := \max\{s,0\}.$$

由于对每一个 $m \le d-2$, $s \mapsto (s)_+^{d-1}$作为一维函数是 C^m，并且由于 $s(t,x)$ 在 (t,x) 上是光滑的，复合函数 $\big(s(t,x)\big)_+^{d-1}$ 对所有 $m \le d-2$ 在 (t,x) 上是 C^m。由此可知， J是C^m，因为它是一个光滑函数与 $\big(s(t,x)\big)_+^{d-1}$ 的乘积。 □

注 2. v 在 (t,x) 中对 $t \in [0,1)$ 为 C^2 这一事实，源于 J 的光滑性，以及 对数映射 的下列性质：*(1)* 对数映射 一致有界，以及 *(2)* 对数映射$_x(x_1)$ 的导数在 $x_1 \to -x$ 处的奇异性在极坐标下得到良好控制。

流匹配常微分方程的存在性与唯一性源于关于时间相关流的经典理论，例如参见（*Marsden* 等人，*2002*，第 *4.1* 节）以及（*Lee*，*2012*年，定理 *9.48*）。

D.1 辅助引理

下面的结果研究仅依赖于径向距离的函数的导数。换句话说，若定义在 M 上的函数 $\phi(x)$ 仅通过 $r(x) = d(x, x_1)$ 依赖于 x，其中 x_1 固定，那么我们就可以通过 r 的导数来控制它的导数。

引理 D.4. 固定 $x_1 \in \mathcal{S}^d$，并定义 $r(x) := d(x, x_1)$。假设 $x \notin$ 割迹集 (x_1)，因此 r 在 x 的某个邻域内光滑，且 $\|$ 梯度 $_x\, r(x)\| = 1$。设 $F : (0, \pi) \to \mathbb{R}$ 为 C^2，并令 $\phi(x) := F(r(x))$。那么

$$\operatorname{grad}_x \phi(x) = F'(r(x))\, \operatorname{grad}_x r(x), \qquad \|\operatorname{grad}_x \phi(x)\| = |F'(r(x))|.$$

此外，

$$\|\nabla \operatorname{grad}\, \phi(x)\|_{\mathrm{op}} \le |F'(r(x))|\, \|\nabla \operatorname{grad}\, r(x)\|_{\mathrm{op}} + |F''(r(x))|.$$

证明。 我们对黎曼梯度应用链式法则。对于任意 $u \in T_x S^d$，微分满足

$$d\phi(x)[u] = d(F \circ r)(x)[u] = F'(r(x))\, dr(x)[u].$$

根据黎曼梯度的定义，

$$\langle \operatorname{grad}_x \phi, u\rangle = F'(r(x))\, \langle \operatorname{grad}_x r, u\rangle, \qquad \forall u \in T_x S^d,$$

因此

$$\operatorname{grad}_x \phi(x) = F'(r(x))\, \operatorname{grad}_x r(x).$$

取范数可得

$$\|\operatorname{grad}_x \phi(x)\| = |F'(r(x))|\,\|\operatorname{grad}_x r(x)\|.$$

Since $x \notin$ 割迹集(x_1)，距离函数满足 $\|$梯度$_x\, r(x)\| = 1$，因此

$$\|\operatorname{grad}_x \phi(x)\| = |F'(r(x))|.$$

接下来，对于任意 $u \in T_x S^d$，

$$\begin{aligned}
\nabla_u \operatorname{grad}\, \phi(x) &= \nabla_u \big(F'(r(x))\, \operatorname{grad}\, r(x)\big) \\
&= F'(r(x))\, \nabla_u \operatorname{grad}\, r(x) + u\, \big(F'(r(x))\big) \operatorname{grad}\, r(x) \\
&= F'(r(x))\, \nabla_u \operatorname{grad}\, r(x) + F''(r(x))\, u(r(x))\, \operatorname{grad}\, r(x) \\
&= F'(r(x))\, \nabla_u \operatorname{grad}\, r(x) + F''(r(x))\, \langle \operatorname{grad}\, r(x), u\rangle\, \operatorname{grad}\, r(x).
\end{aligned}$$

因此，对于 $\|u\| = 1$,

$$\begin{aligned}
\|\nabla_u \operatorname{grad}\, \phi(x)\| &\le |F'(r(x))|\, \|\nabla_u \operatorname{grad}\, r(x)\| + |F''(r(x))|\, |\langle \operatorname{grad}\, r(x), u\rangle|\, \|\operatorname{grad}\, r(x)\| \\
&\le |F'(r(x))|\, \|\nabla_u \operatorname{grad}\, r(x)\| + |F''(r(x))|,
\end{aligned}$$

利用 $\|$梯度 $r(x)\| = 1$ 和 $|\langle$梯度 $r(x)\ u\rangle| \le \|u\| = 1$。对 $\|u\| = 1$ 取上确界可得

$$\|\nabla \operatorname{grad}\, \phi(x)\|_{\mathrm{op}} \le |F'(r(x))|\, \|\nabla \operatorname{grad}\, r(x)\|_{\mathrm{op}} + |F''(r(x))|.$$

□

以下结果控制某些“奇异函数”的条件期望，这将用于建立正则性。

引理 D.5 (期望 $\frac{1}{\sin^a u}$。) 设 $d \ge 3$ 和 $a = 1,\ 2$。令 $r = d(x, x_1)$ 并设$u = r/(1-t) \in (0, \pi)$。我们有

$$\mathbb{E}[\frac{1}{\sin^a u} \mid X_t = x] \le 2\frac{M_1}{m_1},$$

其中 $0 < m_1 \le p_1 \le M_1$。

证明。我们有

$$\begin{aligned}
\mathbb{E}\big[\frac{1}{\sin^a u} \mid X_t = x\big] &= \int_{S^d} \frac{1}{\sin^a u}\, p_t(x_1 \mid x)\, dV_g(x_1) \\
&\le \frac{M_1}{m_1 \operatorname{Vol}(S^d)} \int_{S^d} \frac{1}{\sin^a u}\, J_t(x \mid x_1)\, dV_g(x_1).
\end{aligned}$$

使用 $r = (1-t)u$，使得$\frac{1}{1-t} dr = du$,

$$\begin{aligned}
&\int_{S^d} \frac{1}{\sin^a u}\, J_t(x \mid x_1)\, dV_g(x_1) \\
&\quad = \int_{S^{d-1}} \int_0^\pi \frac{1}{\sin^a u}\, J_t(x \mid x_1)\, (\sin r)^{d-1} dr d\omega \\
&\quad = \operatorname{Vol}(S^{d-1}) \int_0^\pi \frac{1}{\sin^a u} \sin(u)^{d-1}\, du
\end{aligned}$$

$$= \operatorname{Vol}(S^{d-1}) \int_0^\pi \sin(u)^{d-1-a}\, du.$$

因此

$$\begin{aligned}
\mathbb{E}\big[\frac{1}{\sin^a u} \mid X_t = x\big] &= \int_{S^d} \frac{1}{\sin^a u}\, p_t(x_1 \mid x)\, dV_g(x_1) \\
&\le \frac{M_1 \operatorname{Vol}(S^{d-1})}{m_1 \operatorname{Vol}(S^d)} \int_0^\pi \sin(u)^{d-1-a}\, du. \\
&= \frac{M_1 \operatorname{Vol}(S^{d-1})}{m_1 \operatorname{Vol}(S^d) \operatorname{Vol}(S^{d-1-a})} \operatorname{Vol}(S^{d-1-a}) \int_0^\pi \sin(u)^{d-1-a}\, du \\
&= \frac{M_1 \operatorname{Vol}(S^{d-1})}{m_1 \operatorname{Vol}(S^d) \operatorname{Vol}(S^{d-1-a})} \operatorname{Vol}(S^{d-a}).
\end{aligned}$$

回顾

$$\operatorname{Vol}(S^n) = \frac{2\pi^{(n+1)/2}}{\Gamma((n+1)/2)}.$$

因此

$$\begin{aligned}
\frac{\operatorname{Vol}(S^{d-1})}{\operatorname{Vol}(S^d)} &= \frac{\frac{2\pi^{d/2}}{\Gamma(d/2)}}{\frac{2\pi^{(d+1)/2}}{\Gamma((d+1)/2)}} = \pi^{-1/2} \frac{\Gamma((d+1)/2)}{\Gamma(d/2)}, \\
\frac{\operatorname{Vol}(S^{d-a})}{\operatorname{Vol}(S^{d-1-a})} &= \frac{\frac{2\pi^{(d-a+1)/2}}{\Gamma((d-a+1)/2)}}{\frac{2\pi^{(d-a)/2}}{\Gamma((d-a)/2)}} = \pi^{1/2} \frac{\Gamma((d-a)/2)}{\Gamma((d-a+1)/2)}.
\end{aligned}$$

由Kershaw不等式，对 $d \ge 3$,

$$\frac{\Gamma((d+1)/2)}{\Gamma(d/2)} \le (d/2 - \frac{1}{2} + \sqrt{\frac{3}{4}})^{\frac{1}{2}},$$

and

$$\frac{\Gamma(d/2)}{\Gamma((d-1)/2)} \ge (\frac{d}{2} - 1 + \frac{1}{2})^{\frac{1}{2}},$$

$$\frac{\Gamma((d-a+1)/2)}{\Gamma((d-a)/2)} \ge (\frac{d-a-1}{2} + \frac{1}{2})^{\frac{1}{2}}.$$

综上,

$$\begin{aligned}
\frac{\operatorname{Vol}(S^{d-1}) \operatorname{Vol}(S^{d-a})}{\operatorname{Vol}(S^d) \operatorname{Vol}(S^{d-1-a})} &= \frac{\Gamma((d+1)/2)\Gamma((d-a)/2)}{\Gamma(d/2)\Gamma((d-a+1)/2)} \le (\frac{d/2 - \frac{1}{2} + \sqrt{\frac{3}{4}}}{\frac{d-a-1}{2} + \frac{1}{2}})^{\frac{1}{2}} \\
&= (\frac{d-a+a-1+2\sqrt{\frac{3}{4}}}{d-a})^{\frac{1}{2}} \le (1 + \frac{a-1+2\sqrt{\frac{3}{4}}}{d-a})^{\frac{1}{2}}.
\end{aligned}$$

对 $a = 1, 2$，我们有$(1 + \frac{a-1+2\sqrt{\frac{3}{4}}}{d-a})^{\frac{1}{2}} \le 2$。 □

以下结果（引理 D.6、D.7、D.8 以及 D.9）给出了出现在 $v(t,x)$的导数公式中的基本构件的界。对于以下所有结果，我们假设 $0 < m_1 \le p_1(x) \le M_1$。

引理 D.6 （梯度$_x$ 对数 $p_t(X_1 \mid x)$的矩界）。 我们有

$$\| \operatorname{grad}_x \log J_t(x \mid x_1)\| \leq \frac{2(d-1)}{(1-t)\sin u}.$$

此外，

$$\mathbb{E}[\| \operatorname{grad}_x \log p_t(X_1 \mid x)\| \mid X_t = x] \leq \frac{8(d-1)}{(1-t)} \frac{M_1}{m_1},$$
$$\mathbb{E}[\| \operatorname{grad}_x \log J_t(x \mid X_1)\|^2 \mid X_t = x] \leq \frac{8(d-1)^2}{(1-t)^2} \frac{M_1}{m_1},$$
$$\mathbb{E}[\| \operatorname{grad}_x \log p_t(X_1 \mid x)\|^2 \mid X_t = x] \leq \frac{32(d-1)^2}{(1-t)^2} \frac{M_1}{m_1}.$$

证明。 我们首先对 $\|$梯度$_x$对数映射 $J_t(x \mid x_1)\|$进行界定。令 $r = d(x,\ x_1)$，并设 $u = r/(1-t) \in (0,\ \pi)$。在集合 $r < (1-t)\pi$ 上，

$$\frac{\partial}{\partial r} \log J_t(r) = (d-1)\Big(\frac{1}{1-t}\cot\frac{r}{1-t} - \cot r\Big) = (d-1)\Big(\frac{1}{1-t}\cot u - \cot((1-t)u)\Big).$$

根据引理 D.4，

$$\| \operatorname{grad}_x \log J_t(x \mid x_1)\| = \Big|\frac{\partial}{\partial r} \log J_t(r)\Big| = (d-1)\Big|\frac{1}{1-t}\cot u - \cot((1-t)u)\Big|.$$

对于 $u \in (0,\ \pi)$，我们有 $|\cot u| \leq 1/\sin u$。由于 $\sin((1-t)u) \geq (1-t)\sin u$，因此

$$|\cot((1-t)u)| \leq \frac{1}{\sin((1-t)u)} \leq \frac{1}{(1-t)\sin u}.$$

因此，对于所有 $u \in (0,\ \pi)$，我们可以给出界

$$\Big|\frac{1}{1-t}\cot u - \cot((1-t)u)\Big| \leq \frac{1}{1-t}|\cot u| + |\cot((1-t)u)| \leq \frac{1}{(1-t)\sin u} + \frac{1}{(1-t)\sin u}$$
$$= \frac{2}{(1-t)\sin u}.$$

So

$$\| \operatorname{grad}_x \log J_t(x \mid x_1)\| \leq \frac{2(d-1)}{(1-t)\sin u}.$$

现在我们用 $\mathbb{E}\|$梯度$_x$对数映射 $J_t\|$来界定 $\mathbb{E}\|$梯度$_x$对数映射 $p_t\|$。令

$$Z_t(x) := \int_{S^d} p_1(z) J_t(x \mid z)\, dV_g(z),$$

so

$$\log p_t(x_1 \mid x) = \log p_1(x_1) + \log J_t(x \mid x_1) - \log Z_t(x).$$

由于 p_1 不依赖于 x，

$$\operatorname{grad}_x \log p_t(x_1 \mid x) = \operatorname{grad}_x \log J_t(x \mid x_1) - \operatorname{grad}_x \log Z_t(x).$$

此外,

$$\operatorname{grad}_x Z_t(x) = \int_{S^d} p_1(z) \operatorname{grad}_x J_t(x \mid z)\, dV_g(z) = \int_{S^d} p_1(z) J_t(x \mid z) \operatorname{grad}_x \log J_t(x \mid z)\, dV_g(z),$$

因此

$$\operatorname{grad}_x \log Z_t(x) = \int_{S^d} \operatorname{grad}_x \log J_t(x \mid z)\, p_t(z \mid x)\, dV_g(z) = \mathbb{E}\big[\operatorname{grad}_x \log J_t(x \mid X_1) \mid X_t = x\big].$$

因此

$$\operatorname{grad}_x \log p_t(x_1 \mid x) = \operatorname{grad}_x \log J_t(x \mid x_1) - \mathbb{E}\big[\operatorname{grad}_x \log J_t(x \mid X_1) \mid X_t = x\big],$$

并由三角不等式,

$$\mathbb{E}\big[\|\operatorname{grad}_x \log p_t(X_1 \mid x)\| \mid X_t = x\big] \le 2\,\mathbb{E}\big[\|\operatorname{grad}_x \log J_t(x \mid X_1)\| \mid X_t = x\big] \le \frac{8(d-1)}{(1-t)} \frac{M_1}{m_1},$$

其中我们用到了引理D.5。

对于二阶矩，我们有

$$\mathbb{E}[\|\operatorname{grad}_x \log p_t(X_1 \mid x)\|^2 \mid X_t = x] \le 4\,\mathbb{E}[\|\operatorname{grad}_x \log J_t(x \mid X_1)\|^2 \mid X_t = x].$$

回顾 $\|$ 梯度$_x$对数映射 $J_t(x \mid x_1)\| \le$ $^2\frac{(d-1)\,(1-t)}{}$正弦 u，利用引理D.5,

$$\mathbb{E}[\|\operatorname{grad}_x \log J_t(x \mid X_1)\|^2 \mid X_t = x] \le \frac{4(d-1)^2}{(1-t)^2} \mathbb{E}\Big[\frac{1}{\sin^2 u} \mid X_t = x\Big] \le \frac{8(d-1)^2}{(1-t)^2} \frac{M_1}{m_1}.$$

因此

$$\mathbb{E}[\|\operatorname{grad}_x \log p_t(X_1 \mid x)\|^2 \mid X_t = x] \le \frac{32(d-1)^2}{(1-t)^2} \frac{M_1}{m_1}.$$

□

引理 D.7 （ $\partial_t \log p_t(X_1 \mid x)$ 的矩界。）。 我们有

$$\mathbb{E}\Big[\Big|\frac{\partial}{\partial t} \log p_t(X_1 \mid x)\Big| \mid X_t = x\Big] \le \frac{2}{1-t} + 4\pi \frac{d-1}{1-t} \frac{M_1}{m_1},$$
$$\mathbb{E}\Big[\Big|\frac{\partial}{\partial t} \log p_t(X_1 \mid x)\Big|^2 \mid X_t = x\Big] \le \frac{8}{(1-t)^2} + \frac{16\pi^2(d-1)^2}{(1-t)^2} \frac{M_1}{m_1}.$$

证明。设

$$Z_t(x) = \int_{S^d} p_1(z) J_t(x \mid z)\, dV_g(z),$$

因此我们有对数映射 $p_t(x_1 \mid x)$ = 对数映射 $p_1(x_1)$ + 对数映射 $J_t(x \mid x_1)$ − 对数映射 $Z_t(x)$。由于 p_1 不依赖于 t,

$$\frac{\partial}{\partial t} \log p_t(x_1 \mid x) = \frac{\partial}{\partial t} \log J_t(x \mid x_1) - \frac{\partial}{\partial t} \log Z_t(x).$$

此外，对于 $d \geq 2$，函数 $\partial_t J_t(x \mid z)$ 关于 $p_1(z)dV_g(z)$ 可积，且映射 $t \mapsto Z_t(x)$ 是 C^1，并且

$$\frac{\partial}{\partial t} Z_t(x) = \int_{S^d} p_1(z) \frac{\partial}{\partial t} J_t(x \mid z)\, dV_g(z) = \int_{S^d} p_1(z) J_t(x \mid z) \frac{\partial}{\partial t} \log J_t(x \mid z)\, dV_g(z).$$

两边除以 $Z_t(x)$，得到

$$\frac{\partial}{\partial t} \log Z_t(x) = \int_{S^d} \frac{\partial}{\partial t} \log J_t(x \mid z)\, p_t(z \mid x)\, dV_g(z) = \mathbb{E}\Big[\frac{\partial}{\partial t} \log J_t(x \mid X_1) \mid X_t = x\Big].$$

因此

$$\frac{\partial}{\partial t} \log p_t(x_1 \mid x) = \frac{\partial}{\partial t} \log J_t(x \mid x_1) - \mathbb{E}\Big[\frac{\partial}{\partial t} \log J_t(x \mid X_1) \mid X_t = x\Big],$$

再根据三角不等式，

$$\begin{aligned}
\mathbb{E}\Big[\Big|\frac{\partial}{\partial t} \log p_t(X_1 \mid x)\Big| \mid X_t = x\Big] &\leq 2\, \mathbb{E}\Big[\Big|\frac{\partial}{\partial t} \log J_t(x \mid X_1)\Big| \mid X_t = x\Big], \\
\mathbb{E}\Big[\Big|\frac{\partial}{\partial t} \log p_t(X_1 \mid x)\Big|^2 \mid X_t = x\Big] &\leq 4\, \mathbb{E}\Big[\Big|\frac{\partial}{\partial t} \log J_t(x \mid X_1)\Big|^2 \mid X_t = x\Big].
\end{aligned}$$

我们显式计算 ∂_t 对数映射 J_t，并对其条件期望进行界定。对于 $r = d(x, x_1) < (1-t)\pi$，

$$\log J_t(x \mid x_1) = -\log(1-t) + (d-1)\Big(\log \sin(r/(1-t)) - \log \sin r\Big).$$

固定 r，并对 t 求导，得到

$$\frac{\partial}{\partial t} \log J_t(x \mid x_1) = \frac{1}{1-t} + (d-1)\frac{r}{(1-t)^2} \cot(r/(1-t)).$$

记 $u = r/(1-t)$ 为 $\in (0, \pi)$。则 $r = (1-t)u$，且

$$\begin{aligned}
\Big|\frac{\partial}{\partial t} \log J_t(x \mid x_1)\Big| &\leq \frac{1}{1-t} + \frac{d-1}{1-t} |u \cot u|, \\
\Big|\frac{\partial}{\partial t} \log J_t(x \mid x_1)\Big|^2 &\leq 2\frac{1}{(1-t)^2} + 2\frac{(d-1)^2}{(1-t)^2} |u \cot u|^2.
\end{aligned}$$

因此

$$\begin{aligned}
\mathbb{E}\Big[\Big|\frac{\partial}{\partial t} \log J_t(x \mid X_1)\Big| \mid X_t = x\Big] &\leq \frac{1}{1-t} + \pi \frac{d-1}{1-t} \frac{2M_1}{m_1}, \\
\mathbb{E}\Big[\Big|\frac{\partial}{\partial t} \log J_t(x \mid X_1)\Big|^2 \mid X_t = x\Big] &\leq \frac{2}{(1-t)^2} + \frac{2\pi^2(d-1)^2}{(1-t)^2} \frac{2M_1}{m_1}.
\end{aligned}$$

其中我们使用了 $|u$ 余切 $u| \leq {}^{\pi}_{\text{正弦 } u}$。因此，

$$\begin{aligned}
\mathbb{E}\Big[\Big|\frac{\partial}{\partial t} \log p_t(X_1 \mid x)\Big| \mid X_t = x\Big] &\leq \frac{2}{1-t} + 4\pi \frac{d-1}{1-t} \frac{M_1}{m_1}, \\
\mathbb{E}\Big[\Big|\frac{\partial}{\partial t} \log p_t(X_1 \mid x)\Big|^2 \mid X_t = x\Big] &\leq \frac{8}{(1-t)^2} + \frac{16\pi^2(d-1)^2}{(1-t)^2} \frac{M_1}{m_1}.
\end{aligned}$$

□

引理 D.8。我们有

$$\mathbb{E}\big[\|\nabla_x^2 \log p_t(X_1 \mid x)\|_{\mathrm{op}} \mid X_t = x\big] \le \frac{64(d-1)^2}{(1-t)^2}\frac{M_1}{m_1},$$

证明。 回顾可知

$$\operatorname{grad}_x \log p_t(x_1 \mid x) = \operatorname{grad}_x \log J_t(x \mid x_1) - \operatorname{grad}_x \log Z_t(x),$$

再次求导，得到

$$\nabla \operatorname{grad}_x \log p_t(x_1 \mid x) = \nabla \operatorname{grad}_x \log J_t(x \mid x_1) - \nabla \operatorname{grad}_x \log Z_t(x),$$

回顾

$$\operatorname{grad}_x \log Z_t(x) = \mathbb{E}\big[\operatorname{grad}_x \log J_t(x \mid X_1) \mid X_t = x\big].$$

因此

$$\begin{aligned}
&\nabla_u \operatorname{grad}_x \log Z_t(x) = \nabla_u \mathbb{E}\big[\operatorname{grad}_x \log J_t(x \mid X_1) \mid X_t = x\big] \\
=&\mathbb{E}\big[\nabla_u \operatorname{grad}_x \log J_t(x \mid X_1) \mid X_t = x\big] + \mathbb{E}\big[\operatorname{grad}_x \log J_t(x \mid X_1) \nabla_u \log p_t(x_1 \mid x) \mid X_t = x\big],
\end{aligned}$$

并且我们可以给出上界

$$\begin{aligned}
&\|\nabla \operatorname{grad}_x \log Z_t(x)\|_{\mathrm{op}} \\
&\le \mathbb{E}\big[\|\nabla^2 \log J_t(x \mid X_1)\|_{\mathrm{op}} \mid X_t = x\big] \\
&\quad + \sqrt{\mathbb{E}\big[\|\operatorname{grad}_x \log J_t(x \mid X_1)\|^2 \mid X_t = x\big]\mathbb{E}\big[\|\nabla \log p_t(x_1 \mid x)\|^2 \mid X_t = x\big]} \\
&\le \mathbb{E}\big[\|\nabla^2 \log J_t(x \mid X_1)\|_{\mathrm{op}} \mid X_t = x\big] + \frac{16(d-1)^2 M_1}{(1-t)^2 m_1}.
\end{aligned}$$

为了控制 $\|\nabla^2$ 对数映射 $J_t(x \mid X_1)\|_{\mathrm{op}}$，我们考虑

$$\begin{aligned}
\frac{\partial^2}{\partial r^2}\log J_t(x \mid X_1) &= \frac{\partial}{\partial r}(d-1)\Big(\frac{1}{1-t}\cot\frac{r}{1-t} - \cot r\Big) = (d-1)\Big(-\frac{1}{(1-t)^2}\frac{1}{\sin^2\frac{r}{1-t}} + \frac{1}{\sin^2 r}\Big) \\
&= (d-1)\Big(\frac{1}{\sin^2(1-t)u} - \frac{1}{(1-t)^2}\frac{1}{\sin^2 u}\Big).
\end{aligned}$$

应用引理D.4 ， 取 $\phi =$ 对数映射 $J_t(r)$， 可得

$$\begin{aligned}
\|\nabla \operatorname{grad}\, \log J_t(r)\|_{\mathrm{op}} &\le |\log J_t'(r(x))| \|\nabla \operatorname{grad}\, r(x)\|_{\mathrm{op}} + |\log J_t''(r(x))| \\
&\le |\log J_t'(r(x))| \cot r + |\log J_t''(r(x))|,
\end{aligned}$$

其中 $\|\nabla$ 梯度 $r(x)\|_{\mathrm{op}} \le |$余切 $r|$ 由Lee (2018， 命题 11.3)现在利用

$$\begin{aligned}
\frac{\partial}{\partial r}\log J_t(x \mid X_1) &= (d-1)\Big(\frac{1}{1-t}\cot\frac{r}{1-t} - \cot r\Big), \\
\frac{\partial^2}{\partial r^2}\log J_t(x \mid X_1) &= (d-1)\Big(\frac{1}{\sin^2(1-t)u} - \frac{1}{(1-t)^2}\frac{1}{\sin^2 u}\Big),
\end{aligned}$$

我们得到

$$\|\nabla^2 \log J_t(x \mid X_1)\|_{\mathrm{op}} \le \frac{4(d-1)}{(1-t)^2 \sin^2 u},$$

因此

$$\mathbb{E}[\|\nabla^2 \log J_t(x \mid X_1)\| \mid X_t = x] \le \frac{16(d-1)}{(1-t)^2}\frac{M_1}{m_1}.$$

现在我们对 $\mathbb{E}[\|\nabla_x^2$ 对数映射 $p_t(X_1 \mid x)\|_{\mathrm{op}} \mid X_t = x]$进行界定。回顾

$$\begin{aligned}
\nabla \operatorname{grad}_x \log p_t(x_1 \mid x) &= \nabla \operatorname{grad}_x \log J_t(x \mid x_1) - \nabla \operatorname{grad}_x \log Z_t(x), \\
\|\nabla \operatorname{grad}_x \log Z_t(x)\|_{\mathrm{op}} &\le \mathbb{E}\big[\|\nabla^2 \log J_t(x \mid X_1)\|_{\mathrm{op}} \mid X_t = x\big] + \frac{16(d-1)^2 M_1}{(1-t)^2 m_1}.
\end{aligned}$$

由三角不等式，我们得到

$$\mathbb{E}\big[\|\nabla_x^2 \log p_t(X_1 \mid x)\|_{\mathrm{op}} \mid X_t = x\big] \le 2\,\mathbb{E}\big[\|\nabla_x^2 \log J_t(x \mid X_1)\|_{\mathrm{op}} \mid X_t = x\big] + \frac{16(d-1)^2}{(1-t)^2}\frac{M_1}{m_1}.$$

最后，利用该界

$$\mathbb{E}\big[\|\nabla_x^2 \log J_t(x \mid X_1)\|_{\mathrm{op}} \mid X_t = x\big] \le \frac{16(d-1)}{(1-t)^2}\frac{M_1}{m_1},$$

我们得出结论

$$\mathbb{E}\big[\|\nabla_x^2 \log p_t(X_1 \mid x)\|_{\mathrm{op}} \mid X_t = x\big] \le \frac{32(d-1)}{(1-t)^2}\frac{M_1}{m_1} + \frac{16(d-1)^2}{(1-t)^2}\frac{M_1}{m_1} < \frac{64(d-1)^2}{(1-t)^2}\frac{M_1}{m_1}.$$

□

引理 D.9。我们有

$$\mathbb{E}[\|\partial_t \operatorname{grad}_x \log p_t(X_1 \mid x)\| \mid X_t = x] \le \frac{16\pi^2(d-1)^2}{(1-t)^2}\frac{M_1}{m_1}.$$

证明。我们有

$$\begin{aligned}
\operatorname{grad}_x \log p_t(x_1 \mid x) &= \operatorname{grad}_x \log J_t(x \mid x_1) - \operatorname{grad}_x \log Z_t(x) \\
&= \operatorname{grad}_x \log J_t(x \mid x_1) - \mathbb{E}\big[\operatorname{grad}_x \log J_t(x \mid X_1) \mid X_t = x\big],
\end{aligned}$$

因此我们有

$$\begin{aligned}
\|\partial_t \operatorname{grad}_x \log p_t(x_1 \mid x)\| \le{}& \|\partial_t \operatorname{grad}_x \log J_t(x \mid x_1)\| + \mathbb{E}\big[\|\partial_t \operatorname{grad}_x \log J_t(x \mid X_1)\| \mid X_t = x\big] \\
&+ \mathbb{E}[\|\operatorname{grad}_x \log J_t(x \mid X_1)\|\,|\partial_t \log p_t(X_1 \mid x)| \mid X_t = x].
\end{aligned}$$

在固定的 r 下直接微分，得到

$$\begin{aligned}
\|\partial_t \operatorname{grad}_x \log J_t(x \mid x_1)\| &= \|\partial_t \frac{\partial}{\partial r} \log J_t(r) \operatorname{grad} r(x)\| = (d-1)\Big\|\partial_t \frac{1}{1-t}\cot u - \cot((1-t)u)\Big\| \\
&= (d-1)\Big\|\frac{1}{(1-t)^2}\cot u - \frac{u}{\sin^2(1-t)u}\Big\|
\end{aligned}$$

$$\leq \frac{d-1}{(1-t)^2}\frac{2\pi}{\sin^2 u}.$$

取期望，并对 $a=1,2$ 使用 $\mathbb{E}[1/\sin^a u] \leq 2M_1/m_1$，

$$\mathbb{E}[\|\partial_t \operatorname{grad}_x \log J_t(x \mid X_1)\| \mid X_t = x] \leq \frac{4\pi(d-1)}{(1-t)^2}\frac{M_1}{m_1},$$

回顾一下，

$$\begin{aligned}\mathbb{E}\Big[\Big|\partial_t \log p_t(X_1 \mid x)\Big|^2 \mid X_t = x\Big] &\leq 4\mathbb{E}\Big[\Big|\partial_t \log J_t\Big|^2 \mid X_t = x\Big] \\ &\leq 4\mathbb{E}\Big[\frac{(d-1)^2\pi^2}{(1-t)^2}\frac{1}{\sin^2 u} \mid X_t = x\Big] \leq \frac{8\pi^2(d-1)^2}{(1-t)^2}\frac{M_1}{m_1}.\end{aligned}$$

结合 $\|$梯度$_x$对数映射 $J_t(x \mid x_1)\| \leq {}^{2}\!/_{(d-1)\,(1-t)\,\text{正弦 } u}$，由柯西-施瓦茨不等式，

$$\begin{aligned}&\mathbb{E}[\|\operatorname{grad}_x \log J_t(x \mid X_1)\|\,|\partial_t \log p_t(X_1 \mid x)| \mid X_t = x] \\ &\leq \Big(\mathbb{E}[\|\operatorname{grad}_x \log J_t(x \mid X_1)\|^2 \mid X_t = x]\Big)^{1/2}\Big(\mathbb{E}[|\partial_t \log p_t(X_1 \mid x)|^2 \mid X_t = x]\Big)^{1/2} \\ &\leq \sqrt{\frac{8\pi(d-1)^2}{(1-t)^2}\frac{M_1}{m_1}\frac{8\pi^2(d-1)^2}{(1-t)^2}\frac{M_1}{m_1}} \leq \frac{8\pi^2(d-1)^2}{(1-t)^2}\frac{M_1}{m_1}.\end{aligned}$$

Thus

$$\mathbb{E}[\|\partial_t \operatorname{grad}_x \log p_t(X_1 \mid x)\| \mid X_t = x] \leq \frac{16\pi^2(d-1)^2}{(1-t)^2}\frac{M_1}{m_1}.$$

□

D.2 流匹配向量场的正则性

在本节中，我们证明空间导数（引理 D.10）和时间导数（引理D.11） $v(t,x)$ 都是均匀有界的。

引理 D.10。假设 $d \geq 3$ 且 $t \in (0,1)$。设 p_1 是定义在 S^d 上的光滑密度，使得

$$0 < m_1 \leq p_1(z) \leq M_1 < \infty.$$

对于所有单位切向量 w，我们有

$$\|\nabla_w v(t,x)\| \leq \frac{12\pi M_1(d-1)}{m_1(1-t)}.$$

因此，$v(t,x)$ 是 L-利普希茨的，其中 $L = {}^{12\pi M_1}\!/_{(d-1)\,m_1(1-t)}$。

证明。固定 $t<1$ 和 $x \in M$，并设 $w \in T_xM$ 为单位切向量。我们考虑 v 沿 w方向的协变导数：

$$\nabla_w v(t,x) = \frac{1}{1-t}\nabla_w \int_M \operatorname{Log}_x(x_1)\, p_t(x_1|x)\, dV_g(x_1)$$

$$= \frac{1}{1-t}\int_M \nabla_w\big(\operatorname{Log}_x(x_1)\big)\, p_t(x_1|x)\, dV_g(x_1) + \frac{1}{1-t}\int_M \operatorname{Log}_x(x_1)\,\langle \operatorname{grad}_x p_t(x_1|x), w\rangle\, dV_g(x_1).$$

定义

$$P(t,x) := \frac{1}{1-t}\int_{S^d} \operatorname{Log}_x(x_1)\langle \operatorname{grad}_x p_t(x_1 \mid x), w\rangle\, dV_g(x_1),$$
$$G(t,x) := \frac{1}{1-t}\int_{S^d} \|\nabla_w \operatorname{Log}_x(x_1)\| p_t(x_1 \mid x)\, dV_g(x_1).$$

我们证明 $\|P(t,\ x)\| \le \frac{8\pi M_1}{m_1}(d-1)(1-t) \quad \forall \|w\| = 1$。首先，我们将 $\|P(t,\ x)\|$ 写成 $\|$ 梯度 $_x$ 对数映射 $p_t\|$ 的条件矩。利用梯度 $_x p_t = p_t$ 梯度 $_x$ 对数映射 p_t 和 $\|$ 对数映射 $_x(x_1)\| = d(x,\ x_1)$，

$$\begin{aligned}\|P(t,x)\| &\le \frac{1}{1-t}\int_{S^d} \|\operatorname{Log}_x(x_1)\|\, \|\operatorname{grad}_x p_t(x_1 \mid x)\|\, dV_g(x_1)\\ &= \frac{1}{1-t}\int_{S^d} \|\operatorname{Log}_x(x_1)\|\, p_t(x_1 \mid x)\, \|\operatorname{grad}_x \log p_t(x_1 \mid x)\|\, dV_g(x_1)\\ &= \frac{1}{1-t}\mathbb{E}\big[\|\operatorname{Log}_x(X_1)\|\, \|\operatorname{grad}_x \log p_t(X_1 \mid x)\| \mid X_t = x\big].\end{aligned}$$

在 $p_t(\cdot \mid x)$ 的支撑集上，我们有 $d(x,\ X_1) < (1-t)\pi$，因此

$$\|\operatorname{Log}_x(X_1)\| \le (1-t)\pi,$$

so

$$\|P(t,x)\| \le \pi\, \mathbb{E}\big[\|\operatorname{grad}_x \log p_t(X_1 \mid x)\| \mid X_t = x\big].$$

利用引理 D.6,

$$\|P(t,x)\| \le \frac{8\pi M_1(d-1)}{m_1(1-t)}.$$

接下来，我们证明 $G(t,\ x) \le 4\pi \frac{M_1}{m_1}(1-t)$。

注意到 ∇_w 对数映射 $_x$（x_1）可与海森矩阵联系起来。记 $r = d$（$x,\ x_1$）$\in$（0，π），很容易证明（例如，参见 Alimisis 等人（2020，附录 B，证明引理 2））

$$\|\nabla_w \operatorname{Log}_x(x_1)\| \le 1 + |r\cot(r)|.$$

现在我们来界定余切项。在 $p_t(\cdot \mid x)$ 的支撑集上，我们有 $r < (1-t)\pi$。令 $u = r/(1-t) \in (0,\ \pi)$，则 $r = (1-t)u$。利用 $|\cot r| \le 1/\sin r$，以及正弦函数的凹性界 $\sin((1-t)u) \ge (1-t)\sin\ u$,

$$|r\cot r| \le \frac{r}{\sin r} = \frac{(1-t)u}{\sin((1-t)u)} \le \frac{(1-t)u}{(1-t)\sin u} = \frac{u}{\sin u} \le \frac{\pi}{\sin u}.$$

因此对于 $r < (1-t)\pi$,

$$\|\nabla_w \operatorname{Log}_x(x_1)\| \le 1 + \frac{\pi}{\sin u} < \frac{2\pi}{\sin u}.$$

使用与之前相同的方法，

$$\int_{S^d} \|\nabla_w \operatorname{Log}_x(x_1)\|\, p_t(x_1 \mid x)\, dV_g(x_1) \le 4\pi \frac{M_1}{m_1}.$$

我们得出结论。两边乘以 $(1-t)^{-1}$，得到

$$G(t,x) \le 4\pi \frac{M_1}{m_1(1-t)}.$$

□

Lem引理 D.11。 固定 $d \ge 3$ 和 $t \in (0,1)$。设 p_1 是定义在 S^d 上的光滑密度，满足

$$0 < m_1 \le p_1(z) \le M_1 < \infty.$$

For

$$v(t,x) = \frac{1}{1-t} \int_{S^d} \operatorname{Log}_x(x_1)\, p_t(x_1 \mid x)\, dV_g(x_1),$$

对于每个 $x \in S^d$，我们都有，

$$\Big\| \frac{d}{dt} v(t,x) \Big\| \le \frac{3\pi}{1-t} + \frac{4\pi^2 d}{1-t} \frac{M_1}{m_1} \le \frac{8\pi^2 d}{1-t} \frac{M_1}{m_1}.$$

证明。 注意，在 $p_t(\cdot \mid x)$ 的支撑集上，我们有 $d(x,x_1) < (1-t)\pi$，因此

$$\|\operatorname{Log}_x(x_1)\| = d(x,x_1) \le (1-t)\pi.$$

因此

$$\|v(t,x)\| \le \frac{1}{1-t} \int_{S^d} \|\operatorname{Log}_x(x_1)\|\, p_t(x_1 \mid x)\, dV_g(x_1) \le \frac{1}{1-t}(1-t)\pi = \pi.$$

我们对 $v(t,x)$ 关于 t 求导。由于 $x(x_1)$ 的对数映射不依赖于 t，

$$\begin{aligned} \frac{d}{dt} v(t,x) &= \frac{d}{dt}\Big(\frac{1}{1-t}\Big) \int_{S^d} \operatorname{Log}_x(x_1)\, p_t(x_1 \mid x)\, dV_g(x_1) + \frac{1}{1-t} \int_{S^d} \operatorname{Log}_x(x_1)\, \frac{\partial}{\partial t} p_t(x_1 \mid x)\, dV_g(x_1) \\ &= \frac{1}{(1-t)^2} \int_{S^d} \operatorname{Log}_x(x_1)\, p_t(x_1 \mid x)\, dV_g(x_1) + \frac{1}{1-t} \int_{S^d} \operatorname{Log}_x(x_1)\, \frac{\partial}{\partial t} p_t(x_1 \mid x)\, dV_g(x_1). \end{aligned}$$

使用

$$\int_{S^d} \operatorname{Log}_x(x_1)\, p_t(x_1 \mid x)\, dV_g(x_1) = (1-t)\, v(t,x),$$

我们有

$$\frac{d}{dt} v(t,x) = \frac{1}{1-t} v(t,x) + \frac{1}{1-t} \int_{S^d} \operatorname{Log}_x(x_1)\, \frac{\partial}{\partial t} p_t(x_1 \mid x)\, dV_g(x_1).$$

由于 $\int_{S^d} p_t(x_1 \mid x)\, dV_g(x_1) = 1$，我们有

$$\int_{S^d} \frac{\partial}{\partial t} p_t(x_1 \mid x)\, dV_g(x_1) = 0.$$

此外，在 $p_t(x_1 \mid x) > 0$ 的任意位置，

$$\frac{\partial}{\partial t} p_t(x_1 \mid x) = p_t(x_1 \mid x) \frac{\partial}{\partial t} \log p_t(x_1 \mid x).$$

因此

$$\begin{aligned}\Big\| \int_{S^d} \operatorname{Log}_x(x_1) \frac{\partial}{\partial t} p_t(x_1 \mid x)\, dV_g(x_1) \Big\| &\le \int_{S^d} \| \operatorname{Log}_x(x_1) \|\, p_t(x_1 \mid x) \Big| \frac{\partial}{\partial t} \log p_t(x_1 \mid x) \Big|\, dV_g(x_1) \\ &\le (1-t)\pi\, \mathbb{E}\Big[\Big| \frac{\partial}{\partial t} \log p_t(X_1 \mid x) \Big| \mid X_t = x \Big].\end{aligned}$$

因此

$$\Big\| \frac{d}{dt} v(t,x) \Big\| \le \frac{\pi}{1-t} + \pi\, \mathbb{E}\Big[\Big| \frac{\partial}{\partial t} \log p_t(X_1 \mid x) \Big| \mid X_t = x \Big].$$

利用引理 D.7，

$$\begin{aligned}\Big\| \frac{d}{dt} v(t,x) \Big\| &\le \frac{\pi}{1-t} + \pi\, \mathbb{E}\Big[\Big| \frac{\partial}{\partial t} \log p_t(X_1 \mid x) \Big| \mid X_t = x \Big] \\ &\le \frac{\pi}{1-t} + 2\pi \Big(\frac{1}{1-t} + \frac{d-1}{1-t} \frac{2\pi M_1}{m_1} \Big) \le \frac{1}{1-t} \Big(3\pi + 4\pi^2 \frac{d M_1}{m_1} \Big).\end{aligned}$$

□

D.3 散度的正则性

在本节中，我们证明散度 div $v(t,\ x)$的梯度（引理 D.12）和时间导数（引理D.13）均有界。

引理 D.12。 假设 $d \ge 3$ 和 $t \in (0,\ 1)$。设 p_1 是定义在 S^d 上的光滑密度，使得

$$0 < m_1 \le p_1(z) \le M_1 < \infty.$$

那么，对于每个 $x \in S^d$，我们都有

$$\| \operatorname{grad}_x \operatorname{div}\, v(t,x) \| \le \frac{128\pi (d-1)^2}{(1-t)^3} \frac{M_1}{m_1}.$$

证明。设

$$v(t,x) = \frac{1}{1-t} \int_{S^d} \operatorname{Log}_x(x_1)\, p_t(x_1 \mid x)\, dV_g(x_1).$$

使用散度 $(fW) = \langle$梯度 $f,\ W\rangle + f$散度 W与 $f = p_t(\cdot \mid x)$ 以及 $W =$对数映射 $_x(\cdot\ )$，

$$\begin{aligned}\operatorname{div}\, v(t,x) &= \frac{1}{1-t} \int_{S^d} \operatorname{div}_x \operatorname{Log}_x(x_1)\, p_t(x_1 \mid x)\, dV_g(x_1) \\ &\quad + \frac{1}{1-t} \int_{S^d} \langle \operatorname{grad}_x p_t(x_1 \mid x), \operatorname{Log}_x(x_1) \rangle\, dV_g(x_1) \\ &= \frac{1}{1-t} \mathbb{E}[\operatorname{div}_x \operatorname{Log}_x(X_1) \mid X_t = x]\end{aligned}$$

$$+\frac{1}{1-t}\mathbb{E}[\langle \operatorname{grad}_x \log p_t(X_1 \mid x), \operatorname{Log}_x(X_1)\rangle \mid X_t = x].$$

沿单位向量 $\xi \in T_x S^d$ 对该恒等式求导，并取范数。我们取梯度，得到（对于光滑被积函数 $F(x, x_1)$），

梯度 $_x \mathbb{E}[F(x, X_1) \mid X_t = x] = \mathbb{E}[$梯度 $_x F(x, X_1) \mid X_t = x] + \mathbb{E}[F(x, X_1)$ 梯度 $_x$ 对数映射 $p_t(X_1 \mid x) \mid X_t = x]$.

应用此公式并利用三角不等式可得

$$\|\operatorname{grad}_x \operatorname{div} v(t,x)\| \le \frac{1}{1-t}\Big(T_1 + T_2 + T_3 + T_4\Big),$$

其中

$$\begin{aligned}
T_1 &= \mathbb{E}[\|\operatorname{grad}_x \operatorname{div}_x \operatorname{Log}_x(X_1)\| \mid X_t = x],\\
T_2 &= \mathbb{E}[|\operatorname{div}_x \operatorname{Log}_x(X_1)| \, \|\operatorname{grad}_x \log p_t(X_1 \mid x)\| \mid X_t = x],\\
T_3 &= \mathbb{E}[\|\operatorname{grad}_x \langle \operatorname{grad}_x \log p_t(X_1 \mid x), \operatorname{Log}_x(X_1)\rangle\| \mid X_t = x],\\
T_4 &= \mathbb{E}[|\langle \operatorname{grad}_x \log p_t(X_1 \mid x), \operatorname{Log}_x(X_1)\rangle| \, \|\operatorname{grad}_x \log p_t(X_1 \mid x)\| \mid X_t = x].
\end{aligned}$$

现在我们计算散度 $_x$ 对数映射 $_x(x_1)$ 及其 x-梯度。设 $r = d(x, x_1) \in (0, \pi)$。注意到

$$\operatorname{Log}_x(x_1) = -\operatorname{grad}_x \frac{1}{2} r^2,$$

因此

$$\operatorname{div}_x \operatorname{Log}_x(x_1) = -\operatorname{div}_x(r \operatorname{grad} r) = -r\Delta r - \langle \operatorname{grad} r, \operatorname{grad} r\rangle = -(d-1) r \cot r - 1,$$

其中我们用到了 $\Delta r = (d-1)\cot r$（例如参见（Lee，2018，定理 11.11））以及 $\|$ 梯度 $r\|^2 = 1$。对 x 求导，利用 r 是一个径向函数，且 $\|$梯度 $r\| = 1$:

$$\operatorname{grad}_x \operatorname{div}_x \operatorname{Log}_x(x_1) = -(d-1)\big(\cot r - r \csc^2 r\big) \operatorname{grad}_x r,$$

这意味着

$$\|\operatorname{grad}_x \operatorname{div}_x \operatorname{Log}_x(x_1)\| = (d-1)\big|\cot r - r \csc^2 r\big| = (d-1)\big|\cot r - r\frac{1}{\sin^2 r}\big|.$$

我们对 $T_1 = \mathbb{E}[\|$ 梯度$_x$ 散度 $_x$ 对数映射 $_x(X_1)\| \mid X_t = x]$ 进行估计。利用 $r = (1-t)u$ 以及 $\sin((1-t)u) \ge (1-t)\sin u$（$\sin$ 在 $[0, \pi]$ 上的凹性），

$$|\cot r| \le \frac{1}{\sin r} \le \frac{1}{(1-t)\sin u}, \qquad r\frac{1}{\sin^2 r} \le \frac{(1-t)\pi}{\sin^2 r} \le \frac{\pi}{(1-t)\sin^2 u}.$$

因此

$$\|\operatorname{grad}_x \operatorname{div}_x \operatorname{Log}_x(x_1)\| \le (d-1)\Big(\frac{1}{(1-t)\sin u} + \frac{\pi}{(1-t)\sin^2 u}\Big).$$

因此

$$T_1 \le (d-1)\frac{1}{1-t}\frac{M_1}{m_1}(2\pi + 2).$$

我们对 $T_2 = \mathbb{E}[|$ 散度 $_x$ 对数映射 $_x(X_1)|$ ‖梯度 $_x$ 对数 $p_t(X_1 \mid x)\| \mid X_t = x]$ 进行界定。回顾一下，$|$ 散度 $_x$ 对数映射 $_x(x_1)| \le 1 + (d-1)r|$余切 $r|$，因此我们有

$$|\operatorname{div}_x \operatorname{Log}_x(x_1)| \le 1 + (d-1)\pi \frac{1}{\sin r} \le 1 + (d-1)\pi \frac{1}{(1-t)\sin u}.$$

因此，根据柯西-施瓦茨不等式，

$$T_2 \le \Big(\mathbb{E}[|\operatorname{div}_x \operatorname{Log}_x(X_1)|^2 \mid X_t = x]\Big)^{1/2} \Big(\mathbb{E}[\|\operatorname{grad}_x \log p_t(X_1 \mid x)\|^2 \mid X_t = x]\Big)^{1/2}.$$

利用 $(a+b)^2 \le 2a^2 + 2b^2$，

$$\begin{aligned}\mathbb{E}[|\operatorname{div}_x \operatorname{Log}_x(X_1)|^2 \mid X_t = x] &\le 2 + 2(d-1)^2\pi^2 \frac{1}{(1-t)^2} \mathbb{E}\Big[\frac{1}{\sin^2 u} \mid X_t = x\Big] \\ &\le 2 + 2(d-1)^2\pi^2 \frac{1}{(1-t)^2} \frac{2M_1}{m_1}.\end{aligned}$$

另请回忆

$$\mathbb{E}[\|\operatorname{grad}_x \log p_t(X_1 \mid x)\|^2 \mid X_t = x] \le \frac{32(d-1)^2}{(1-t)^2} \frac{M_1}{m_1}.$$

因此我们得到

$$\begin{aligned}T_2 &\le (2 + 2(d-1)^2\pi^2 \frac{1}{(1-t)^2} \frac{2M_1}{m_1})^{\frac{1}{2}} (\frac{32(d-1)^2}{(1-t)^2} \frac{M_1}{m_1})^{\frac{1}{2}} \\ &\le \frac{16\pi M_1}{m_1} \frac{(d-1)^2}{(1-t)^2}.\end{aligned}$$

我们对 $T_3 = \mathbb{E}[\|$ 梯度 $_x\langle$梯度 $_x$ 对数映射 $p_t(X_1 \mid x)$，对数映射 $_x(X_1)\rangle\| \mid X_t = x]$ 进行界定。我们使用乘积法则：记 $g(x) = \langle g_1(x)\ g_2(x)\rangle$。对于 g_1，g_2，我们有（将其视为沿 u 的方向导数，并利用相容性）

$$\langle \operatorname{grad}_x \langle g_1(x), g_2(x)\rangle, u\rangle = \nabla_u \langle g_1(x), g_2(x)\rangle = \langle \nabla_u g_1(x), g_2(x)\rangle + \langle g_1(x), \nabla_u g_2(x)\rangle.$$

因此我们可以写成

$$\begin{aligned}&\|\operatorname{grad}_x \langle \operatorname{grad}_x \log p_t(x_1 \mid x), \operatorname{Log}_x(x_1)\rangle\| \\ =&\|\nabla \operatorname{grad}_x \log p_t(x_1 \mid x)\|_{\mathrm{op}} \|\operatorname{Log}_x(x_1)\| + \|\nabla \operatorname{Log}_x(x_1)\|_{\mathrm{op}} \|\operatorname{grad}_x \log p_t(x_1 \mid x)\|,\end{aligned}$$

so

$$\begin{aligned}T_3 \le\ &\mathbb{E}[\|\nabla \operatorname{Log}_x(X_1)\| \, \|\operatorname{grad}_x \log p_t(X_1 \mid x)\| \mid X_t = x] \\ &+ \mathbb{E}[\|\nabla^2 \log p_t(X_1 \mid x)\| \, \|\operatorname{Log}_x(X_1)\| \mid X_t = x].\end{aligned}$$

在 $p_t(\cdot \mid x)$ 的支撑集上，‖对数映射 $_x(X_1)\| \le (1-t)\pi$，因此

$$\mathbb{E}[\|\nabla^2 \log p_t(X_1 \mid x)\| \, \|\operatorname{Log}_x(X_1)\| \mid X_t = x] \le (1-t)\pi\, \mathbb{E}[\|\nabla^2 \log p_t(X_1 \mid x)\| \mid X_t = x].$$

由引理 D.8，

$$\mathbb{E}[\|\nabla^2 \log p_t(X_1 \mid x)\| \, \|\operatorname{Log}_x(X_1)\| \mid X_t = x] \le \frac{64\pi(d-1)^2}{1-t} \frac{M_1}{m_1}.$$

此外，在 S^d 上，

$$\|\nabla \operatorname{Log}_x(x_1)\| \le 1 + r|\cot r| < \pi \frac{2}{(1-t)\sin u},$$

因此，根据柯西-施瓦茨不等式，

$$\begin{aligned}\mathbb{E}[\|\nabla \operatorname{Log}_x(X_1)\| \,\|\operatorname{grad}_x \log p_t(X_1 \mid x)\| \mid X_t = x] &\le \Big(\frac{32(d-1)^2}{(1-t)^2}\frac{M_1}{m_1}\frac{8\pi^2}{(1-t)^2}\frac{M_1}{m_1}\Big)^{\frac{1}{2}} \\ &\le \frac{16\pi M_1}{m_1}\frac{(d-1)}{(1-t)^2}.\end{aligned}$$

因此我们得到

$$T_3 \le \frac{16\pi M_1}{m_1}\frac{(d-1)}{(1-t)^2} + \frac{64\pi(d-1)^2}{1-t}\frac{M_1}{m_1} \le \frac{64\pi M_1}{m_1}\frac{(d-1)^2}{(1-t)^2}.$$

我们对 $T_4 = \mathbb{E}[|\langle$梯度 $_x$ 对数映射 $p_t(X_1 \mid x)$，对数映射 $_x(X_1)\rangle|$ $\|$梯度 $_x$ 对数映射 $p_t(X_1 \mid x)\| \mid X_t = x]$进行估计。我们有

$$\begin{aligned}T_4 &\le \mathbb{E}[\|\operatorname{Log}_x(X_1)\| \,\|\operatorname{grad}_x \log p_t(X_1 \mid x)\|^2 \mid X_t = x] \\ &\le (1-t)\pi\, \mathbb{E}[\|\operatorname{grad}_x \log p_t(X_1 \mid x)\|^2 \mid X_t = x] \\ &\le (1-t)\pi \cdot \frac{32(d-1)^2}{(1-t)^2}\frac{M_1}{m_1} = \frac{32\pi(d-1)^2}{1-t}\frac{M_1}{m_1}.\end{aligned}$$

综合起来，

$$\begin{aligned}&\|\operatorname{grad}_x \operatorname{div}\, v(t,x)\| \\ \le&\frac{1}{1-t}\Big(T_1 + T_2 + T_3 + T_4\Big) \\ \le&\frac{1}{(1-t)}\frac{M_1}{m_1}\Big((d-1)\frac{1}{1-t}(2\pi+2) + 16\pi\frac{(d-1)^2}{(1-t)^2} + 64\pi\frac{(d-1)^2}{(1-t)^2} + \frac{32\pi(d-1)^2}{1-t}\Big) \\ \le&\frac{128\pi(d-1)^2}{(1-t)^3}\frac{M_1}{m_1}.\end{aligned}$$

□

引理 D.13。 假设 $d \ge 3$ 和 $t \in (0,1)$。设 p_1 是 S^d 上的光滑密度，使得

$$0 < m_1 \le p_1(z) \le M_1 < \infty.$$

那么对于每个 $x \in S^d$，

$$\Big|\frac{d}{dt}\operatorname{div}\, v(t,x)\Big| \le \frac{128\pi^2(d-1)^2}{(1-t)^3}\frac{M_1}{m_1}.$$

证明。 我们先计算散度的时间导数。回顾

$$\operatorname{div}\, v(t,x) = \frac{1}{1-t}\mathbb{E}[\operatorname{div}_x \operatorname{Log}_x(X_1) \mid X_t = x] + \frac{1}{1-t}\mathbb{E}[\langle \operatorname{grad}_x \log p_t(X_1 \mid x), \operatorname{Log}_x(X_1)\rangle \mid X_t = x].$$

设 $F(t, x_1)$ 在 t 上光滑，且在 $p_t(\cdot \mid x)$ 下可积。那么

$$\frac{d}{dt}\mathbb{E}[F(t,X_1) \mid X_t = x] = \mathbb{E}[\partial_t F(t,X_1) \mid X_t = x] + \mathbb{E}[F(t,X_1)\,\partial_t \log p_t(X_1 \mid x) \mid X_t = x].$$

分别将 $F(t,x_1)$ 取为 $=$散度 $_x$ 对数映射 $_x(x_1)$，将 $F(t,x_1)$ 取为 $=\langle$梯度 $_x$ 对数 $p_t(x_1 \mid x)$，对数映射 $_x(x_1)\rangle$，应用此结论，

$$\frac{d}{dt}\mathbb{E}[\operatorname{div}_x \operatorname{Log}_x(X_1) \mid X_t = x] = \mathbb{E}[\operatorname{div}_x \operatorname{Log}_x(X_1)\, \partial_t \log p_t(X_1 \mid x) \mid X_t = x],$$
$$\frac{d}{dt}\mathbb{E}[\langle \operatorname{grad}_x \log p_t(X_1 \mid x), \operatorname{Log}_x(X_1)\rangle \mid X_t = x] = \mathbb{E}[\partial_t \langle \operatorname{grad}_x \log p_t(X_1 \mid x), \operatorname{Log}_x(X_1)\rangle \mid X_t = x]$$
$$+ \mathbb{E}[\langle \operatorname{grad}_x \log p_t(X_1 \mid x), \operatorname{Log}_x(X_1)\rangle\, \partial_t \log p_t(X_1 \mid x) \mid X_t = x].$$

因此，对于固定的 x,

$$\begin{aligned}\frac{d}{dt}\operatorname{div} v(t,x) = \frac{1}{(1-t)^2}\mathbb{E}\big[\operatorname{div}_x \operatorname{Log}_x(X_1) + \langle \operatorname{grad}_x \log p_t(X_1 \mid x), \operatorname{Log}_x(X_1)\rangle \mid X_t = x\big] \\ + \frac{1}{1-t}\mathbb{E}[\operatorname{div}_x \operatorname{Log}_x(X_1)\, \partial_t \log p_t(X_1 \mid x) \mid X_t = x] \\ + \frac{1}{1-t}\mathbb{E}[\partial_t \langle \operatorname{grad}_x \log p_t(X_1 \mid x), \operatorname{Log}_x(X_1)\rangle \mid X_t = x] \\ + \frac{1}{1-t}\mathbb{E}[\langle \operatorname{grad}_x \log p_t(X_1 \mid x), \operatorname{Log}_x(X_1)\rangle\, \partial_t \log p_t(X_1 \mid x) \mid X_t = x].\end{aligned}$$

定义

$$\begin{aligned}T_1 &= \frac{1}{(1-t)^2}\mathbb{E}\big[|\operatorname{div}_x \operatorname{Log}_x(X_1) + \langle \operatorname{grad}_x \log p_t(X_1 \mid x), \operatorname{Log}_x(X_1)\rangle| \mid X_t = x\big],\\ T_2 &= \frac{1}{1-t}\mathbb{E}[|\operatorname{div}_x \operatorname{Log}_x(X_1)\, \partial_t \log p_t(X_1 \mid x)| \mid X_t = x],\\ T_3 &= \frac{1}{1-t}\mathbb{E}[|\partial_t \langle \operatorname{grad}_x \log p_t(X_1 \mid x), \operatorname{Log}_x(X_1)\rangle| \mid X_t = x],\\ T_4 &= \frac{1}{1-t}\mathbb{E}[|\langle \operatorname{grad}_x \log p_t(X_1 \mid x), \operatorname{Log}_x(X_1)\rangle\, \partial_t \log p_t(X_1 \mid x)| \mid X_t = x].\end{aligned}$$

对于 T_1，回顾

$$\begin{aligned}\frac{1}{(1-t)^2}\mathbb{E}\big[\operatorname{div}_x \operatorname{Log}_x(X_1) \mid X_t = x\big] &\le \frac{1}{(1-t)^2}\mathbb{E}\Big[1 + (d-1)\pi \frac{1}{(1-t)\sin u} \mid X_t = x\Big] \\ &\le \pi \frac{(d-1)}{(1-t)^3}\frac{2M_1}{m_1},\end{aligned}$$

and

$$\begin{aligned}&\frac{1}{(1-t)^2}\mathbb{E}\big[\langle \operatorname{grad}_x \log p_t(X_1 \mid x), \operatorname{Log}_x(X_1)\rangle \mid X_t = x\big] \\ \le &\frac{1}{(1-t)^2}(1-t)\pi\mathbb{E}\big[\|\operatorname{grad}_x \log p_t(X_1 \mid x)\| \mid X_t = x\big] \le \frac{8\pi(d-1)}{(1-t)^2}\frac{M_1}{m_1}.\end{aligned}$$

因此

$$T_1 \le \pi \frac{(d-1)}{(1-t)^3}\frac{2M_1}{m_1} + \frac{8\pi(d-1)}{(1-t)^2}\frac{M_1}{m_1} \le \frac{16\pi(d-1)}{(1-t)^3}\frac{M_1}{m_1}.$$

对于 T_2，回顾

$$\mathbb{E}[|\text{散度}_x\text{对数映射}_x(X_1)|^2 \mid X_t = x] \le 2 + 2(d-1)^2\pi^2 \quad \frac{1}{(1-t)^2}\frac{2M_1}{m_1} \le \frac{2}{(d-1)^2}\frac{4\pi^2 M_1}{m_1}\cdot(1-t)$$

因此我们有

$$\begin{aligned}T_2 =&\frac{1}{1-t}\mathbb{E}[|\operatorname{div}_x \operatorname{Log}_x(X_1)\,\partial_t \log p_t(X_1 \mid x)| \mid X_t = x]\\ \le&\frac{1}{1-t}\Big(\mathbb{E}[|\operatorname{div}_x \operatorname{Log}_x(X_1)|^2 \mid X_t = x]\mathbb{E}[|\partial_t \log p_t(X_1 \mid x)|^2 \mid X_t = x]\Big)^{\frac{1}{2}}\\ \le&\frac{1}{1-t}\Big(\frac{(d-1)^2}{(1-t)^2}\frac{4\pi^2 M_1}{m_1}\frac{32\pi^2(d-1)^2}{(1-t)^2}\frac{M_1}{m_1}\Big)^{\frac{1}{2}} \le \frac{16\pi^2(d-1)^2}{(1-t)^3}\frac{M_1}{m_1},\end{aligned}$$

根据引理 D.7,

$$\mathbb{E}\Big[\Big|\frac{\partial}{\partial t}\log p_t(X_1 \mid x)\Big|^2 \mid X_t = x\Big] \le \frac{8}{(1-t)^2} + \frac{16\pi^2(d-1)^2}{(1-t)^2}\frac{M_1}{m_1} \le \frac{32\pi^2(d-1)^2}{(1-t)^2}\frac{M_1}{m_1}.$$

对于 T_3，由于 $\mathbf{Log}_x(x_1)$ 不依赖于 t,

$$\begin{aligned}&|\partial_t\langle \operatorname{grad}_x \log p_t(x_1 \mid x), \operatorname{Log}_x(x_1)\rangle|\\ \le&\|\partial_t \operatorname{grad}_x \log p_t(x_1 \mid x)\|\,\|\operatorname{Log}_x(x_1)\| \le (1-t)\pi\,\|\partial_t \operatorname{grad}_x \log p_t(x_1 \mid x)\|.\end{aligned}$$

利用引理 D.9，我们有

$$\mathbb{E}[\|\partial_t \operatorname{grad}_x \log p_t(X_1 \mid x)\| \mid X_t = x] \le \frac{16\pi^2(d-1)^2}{(1-t)^2}\frac{M_1}{m_1}.$$

因此

$$\begin{aligned}\mathbb{E}[|\partial_t\langle \operatorname{grad}_x \log p_t(X_1 \mid x), \operatorname{Log}_x(X_1)\rangle| \mid X_t = x] \le&(1-t)\pi\cdot\frac{16\pi^2(d-1)^2}{(1-t)^2}\frac{M_1}{m_1}\\ =&\frac{16\pi^3(d-1)^2}{(1-t)}\frac{M_1}{m_1}.\end{aligned}$$

最后，我们对 T_4 进行估计。类似地，根据柯西-施瓦茨不等式以及我们对 T_3 所作的估计,

$$\begin{aligned}&\Big|\mathbb{E}[\langle \operatorname{grad}_x \log p_t(X_1 \mid x), \operatorname{Log}_x(X_1)\rangle\,\partial_t \log p_t(X_1 \mid x) \mid X_t = x]\Big|\\ \le&\Big(\mathbb{E}[|\langle \operatorname{grad}_x \log p_t(X_1 \mid x), \operatorname{Log}_x(X_1)\rangle|^2 \mid X_t = x]\Big)^{1/2}\Big(\mathbb{E}[|\partial_t \log p_t(X_1 \mid x)|^2 \mid X_t = x]\Big)^{1/2}.\end{aligned}$$

我们将第一个因子估计为

$$\begin{aligned}&\mathbb{E}[|\langle \operatorname{grad}_x \log p_t(X_1 \mid x), \operatorname{Log}_x(X_1)\rangle|^2 \mid X_t = x]\\ \le&(1-t)^2\pi^2\mathbb{E}[\|\operatorname{grad}_x \log p_t(X_1 \mid x)\|^2 \mid X_t = x] \le 32\pi^2(d-1)^2\frac{M_1}{m_1}.\end{aligned}$$

因此

$$\begin{aligned}T_4 =&\frac{1}{1-t}\Big|\mathbb{E}[\langle \operatorname{grad}_x \log p_t(X_1 \mid x), \operatorname{Log}_x(X_1)\rangle\,\partial_t \log p_t(X_1 \mid x) \mid X_t = x]\Big|\\ \le&32\pi^2\frac{(d-1)^2}{(1-t)^2}\frac{M_1}{m_1}.\end{aligned}$$

综上,

$$|\frac{d}{dt}\operatorname{div}\, v(t,x)| \le T_1 + T_2 + T_3 + T_4$$

$$\leq \frac{16\pi(d-1)}{(1-t)^3}\frac{M_1}{m_1} + \frac{16\pi^2(d-1)^2}{(1-t)^3}\frac{M_1}{m_1} + \frac{16\pi^3(d-1)^2}{(1-t)^2}\frac{M_1}{m_1} + 32\pi^2\frac{(d-1)^2}{(1-t)^2}\frac{M_1}{m_1}$$
$$\leq \frac{128\pi^2(d-1)^2}{(1-t)^3}\frac{M_1}{m_1}.$$

□

$v(t,\ x)$ 的正则性以及对数映射 p_t

最后，我们在引理 $\|v(t,\ x)\|$ 和 $\mathbb{E}[\|$梯度 对数映射 $p_t(X_t)\|^2]$ 中给出了界 D.14 和 D.15。

引理 D.14. 假设 $d \geq 2$ 和 $t \in (0,1)$。设 p_1 是定义在 S^d 上的光滑密度，满足

$$0 < m_1 \leq p_1(z) \leq M_1 < \infty.$$

对于每个 $x \in S^d$,

$$\|v(t,x)\| \leq \pi.$$

证明。 固定 $x \in S^d$。由 Jensen 不等式及 v的定义,

$$\|v(t,x)\| = \frac{1}{1-t}\Big\|\int_{S^d} \operatorname{Log}_x(x_1)\, p_t(x_1 \mid x)\, dV_g(x_1)\Big\| \leq \frac{1}{1-t}\int_{S^d} \|\operatorname{Log}_x(x_1)\|\, p_t(x_1 \mid x)\, dV_g(x_1).$$

在 $p_t(\cdot \mid x)$的支撑集上，我们有 $d(x,\ x_1) < (1-t)\pi$，因此

$$\|\operatorname{Log}_x(x_1)\| = d(x,x_1) \leq (1-t)\pi.$$

因此

$$\|v(t,x)\| \leq \frac{1}{1-t}\int_{S^d}(1-t)\pi\, p_t(x_1 \mid x)\, dV_g(x_1) = \pi\int_{S^d} p_t(x_1 \mid x)\, dV_g(x_1) = \pi,$$

因为 $p_t(\cdot \mid x)$ 是概率密度。 □

引理 D.15. 假设 $d \geq 3$ 和 $t \in (0,1)$。令 $p_0 \equiv \operatorname{Vol}(S^d)^{-1}$ 为 S^d 上的均匀先验，且令 p_1 为 S^d 上的光滑密度，使得

$$0 < m_1 \leq p_1(z) \leq M_1 < \infty.$$

令 $X_0 \sim p_0$ 和 $X_1 \sim p_1$ 相互独立，并定义测地线插值

$$X_t = \operatorname{Exp}_{X_0}(t \operatorname{Log}_{X_0}(X_1)).$$

令 p_t 表示 X_t 关于 dV_g的边缘密度。则

$$\mathbb{E}\big[\|\operatorname{grad}\log p_t(X_t)\|^2\big] \leq \frac{8(d-1)^2}{(1-t)^2}\frac{M_1}{m_1}.$$

证明。 利用引理 D.2，注意到对数映射 $p_t(x)$ = 对数映射 $Z_t(x)$ − 对数映射体积 (S^d)，我们有

$$\operatorname{grad} \log p_t(x) = \operatorname{grad} \log Z_t(x).$$

取梯度，得到

$$\operatorname{grad} Z_t(x) = \int_{S^d} p_1(x_1) \operatorname{grad}_x J_t(x \mid x_1)\, dV_g(x_1).$$

利用梯度 ${}_x J_t = J_t$ 梯度x对数映射 J_t 以及条件密度的定义

$$p_t(x_1 \mid x) = \frac{p_1(x_1) J_t(x \mid x_1)}{Z_t(x)},$$

我们得到

$$\begin{aligned}\operatorname{grad} \log Z_t(x) = \frac{\operatorname{grad} Z_t(x)}{Z_t(x)} &= \int_{S^d} \operatorname{grad}_x \log J_t(x \mid x_1)\, p_t(x_1 \mid x)\, dV_g(x_1) \\ &= \mathbb{E}\big[\operatorname{grad}_x \log J_t(x \mid X_1) \mid X_t = x\big].\end{aligned}$$

现在我们有

$$\operatorname{grad} \log p_t(X_t) = \mathbb{E}[\operatorname{grad}_x \log J_t(X_t \mid X_1) \mid X_t],$$

这意味着

$$\| \operatorname{grad} \log p_t(X_t)\|^2 \le \mathbb{E}[\| \operatorname{grad}_x \log J_t(X_t \mid X_1)\|^2 \mid X_t].$$

再次取期望可得

$$\mathbb{E}[\| \operatorname{grad} \log p_t(X_t)\|^2] \le \mathbb{E}\Big[\mathbb{E}[\| \operatorname{grad}_x \log J_t(X_t \mid X_1)\|^2 \mid X_t]\Big] \le \frac{8(d-1)^2}{(1-t)^2} \frac{M_1}{m_1}.$$

其中最后一个不等式由引理 D.6得出。

□

D.5 评分正则性的有限性：命题 5.3 的证明

证明。 [命题 5.3] 注意到

$$\mathbb{E}\big[\| \operatorname{grad} \log p_t(X_t)\|^2\big] = \int_M \| \operatorname{grad} \log p_t(x)\|^2 p_t(x)\, dV_g(x) =: I(p_t),$$

也就是说，它恰好是 p_t的（黎曼）Fisher 信息。考虑连续性方程$\partial_t p_t +$ 散度 $(p_t\, v(t,\ \cdot)) = 0$。令 $s_t :=$ 梯度 对数映射 p_t 表示得分函数。

我们首先t 对 I求时间导数。进行求导并使用普通乘积法则 :

$$\begin{aligned}\frac{d}{dt} I(p_t) &= \int_M \partial_t\big(\langle s_t, s_t\rangle\big)\, p_t\, dV_g + \int_M \langle s_t, s_t\rangle\, \partial_t p_t\, dV_g \\ &= 2\int_M \big\langle \partial_t s_t,\ s_t\big\rangle\, p_t\, dV_g + \int_M \|s_t\|^2\, \partial_t p_t\, dV_g.\end{aligned}$$

注意到

$$\begin{aligned}\partial_t \log p_t &= -\frac{1}{p_t}\operatorname{div}(p_t v) = -\operatorname{div} v - \frac{1}{p_t}\langle \operatorname{grad} p_t, v\rangle \\ &= -\operatorname{div} v - \langle \operatorname{grad} \log p_t, v\rangle = -\operatorname{div} v - \langle s_t, v\rangle.\end{aligned}$$

取梯度可得

$$\partial_t s_t = \operatorname{grad}(\partial_t \log p_t) = -\operatorname{grad}(\operatorname{div} v) - \operatorname{grad}\big(\langle s_t, v\rangle\big).$$

We c分别计算 $I(p_t)$ 的时间导数中的各项。对于第一项 ,

$$\begin{aligned}2\int_M \langle \partial_t s_t,\, s_t\rangle\, p_t\, dV_g &= 2\int_M \big\langle -\operatorname{grad}(\operatorname{div} v) - \operatorname{grad}\big(\langle s_t, v\rangle\big),\, s_t\big\rangle\, p_t\, dV_g \\ &= -2\int_M \big\langle \operatorname{grad}(\operatorname{div} v),\, s_t\big\rangle\, p_t\, dV_g - 2\int_M \big\langle \operatorname{grad}\big(\langle s_t, v\rangle\big),\, s_t\big\rangle\, p_t\, dV_g \\ &= -2\int_M \big\langle \operatorname{grad}(\operatorname{div} v),\, s_t\big\rangle\, p_t\, dV_g - 2\int_M \langle \nabla_{s_t} s_t, v\rangle p_t + \langle \nabla_{s_t} v, s_t\rangle p_t\, dV_g.\end{aligned}$$

对于第二项,

$$\begin{aligned}\int_M \|s_t\|^2\, \partial_t p_t\, dV_g &= -\int_M \|s_t\|^2 \operatorname{div}(p_t v)\, dV_g = \int_M \big\langle \operatorname{grad}\|s_t\|^2,\, v\big\rangle\, p_t\, dV_g \\ &= 2\int_M \langle \nabla_v s_t, s_t\rangle\, p_t\, dV_g.\end{aligned}$$

由于海森矩阵是对称的，我们有

$$\big\langle \nabla_v s_t, s_t\big\rangle = \langle \nabla_v \operatorname{grad}\log p_t, s_t\rangle = \big(\operatorname{Hess}\log p_t\big)(v, s_t) = \langle \nabla_{s_t} \operatorname{grad}\log p_t, v\rangle = \big\langle \nabla_{s_t} s_t, v\big\rangle.$$

综上,

$$\frac{d}{dt} I(p_t) = -2\int_M \langle s_t, \nabla_{s_t} v\rangle\, p_t\, dV_g - 2\int_M \langle \operatorname{grad}(\operatorname{div} v), s_t\rangle\, p_t\, dV_g.$$

现在我们推导一个有助于界定 I 的常微分方程。利用柯西–施瓦茨不等式以及逐点界$\|\nabla v(t, x)\|_{\mathrm{op}} \le L^{v,x}_t$和 $\|$梯度 散度 $v(t, x)\| \le L^{\mathrm{div},x}_t$（由假设 2），我们得到

$$\begin{aligned}\frac{d}{dt} I(p_t) &\le 2L^{v,x}_t\, I(p_t) + 2L^{\mathrm{div},x}_t \int_M \|s_t(x)\|\, p_t(x)\, dV_g(x) \\ &\le 2L^{v,x}_t\, I(p_t) + 2L^{\mathrm{div},x}_t \sqrt{I(p_t)}. \qquad (16)\end{aligned}$$

令 $y(t) := \sqrt{I(p_t)}$。由于 $I' = 2yy'$，（16）意味着

$$y'(t) \le L^{v,x}_t\, y(t) + L^{\mathrm{div},x}_t.$$

剩下的工作是求解常微分方程并获得一个有限的上界。定义

$$A(t) := \int_0^t L^{v,x}_s\, ds, \qquad \mu(t) := e^{-A(t)}.$$

那么 μ 是绝对连续的，并且

$$\mu'(t) = -L^{v,x}_t\, e^{-A(t)} = -L^{v,x}_t\, \mu(t) \quad \text{for a.e. } t.$$

两边乘以 $\mu(t)$:

$$\mu(t)\, y'(t) \le \mu(t)\, L_t^{v,x}\, y(t) + \mu(t)\, L_t^{\mathrm{div},x}.$$

利用乘积法则以及 $\mu'(t)$的恒等式,

$$\begin{aligned}
\frac{d}{dt}\big(\mu(t)\, y(t)\big) &= \mu'(t)\, y(t) + \mu(t)\, y'(t) \\
&= -L_t^{v,x}\, \mu(t)\, y(t) + \mu(t)\, y'(t) \\
&\le -L_t^{v,x}\, \mu(t)\, y(t) + \mu(t)\, L_t^{v,x}\, y(t) + \mu(t)\, L_t^{\mathrm{div},x} \\
&= \mu(t)\, L_t^{\mathrm{div},x}.
\end{aligned}$$

因此，对几乎处处的 t,

$$\frac{d}{dt}\big(\mu(t)\, y(t)\big) \le \mu(t)\, L_t^{\mathrm{div},x}.$$

对 $[0,t]$ 积分可得

$$\mu(t)\, y(t) - \mu(0)\, y(0) \le \int_0^t \mu(s)\, L_s^{\mathrm{div},x}\, ds.$$

由于 $\mu(0) = e^{-A(0)} = 1$，我们得到

$$\mu(t)\, y(t) \le y(0) + \int_0^t e^{-A(s)}\, L_s^{\mathrm{div},x}\, ds.$$

将两边乘以 $e^{A(t)}$:

$$\begin{aligned}
y(t) &\le e^{A(t)} \left(y(0) + \int_0^t e^{-A(s)}\, L_s^{\mathrm{div},x}\, ds \right) \\
&= \exp\Big(\int_0^t L_s^{v,x}\, ds \Big) \left(y(0) + \int_0^t L_s^{\mathrm{div},x}\, \exp\Big(- \int_0^s L_r^{v,x}\, dr \Big)\, ds \right).
\end{aligned}$$

因此，对所有 $t < 1$,

$$\sqrt{I(p_t)} \le \exp\Big(\int_0^t L_s^{v,x}\, ds \Big) \left(\sqrt{I(p_0)} + \int_0^t L_s^{\mathrm{div},x}\, \exp\Big(- \int_0^s L_r^{v,x}\, dr \Big)\, ds \right),$$

因此

$$\mathbb{E}\big[\| \operatorname{grad} \log p_t(X_t) \|^2 \big] = I(p_t) \le L_t^{\mathrm{score}}$$

对于某个有限数值 L_t^{score}。 □

E SPD流形正则性结果

在本节中，我们研究带有仿射不变度量的 SPD 流形，记作 SPD（n）。

我们简要讨论本节的正则性分析与超球面的正则性分析之间的相似之处和差异。

• 在超球面上存在割点，因此条件密度中会出现示性函数。关于一般 Hadamard 流形上的条件密度函数公式，见引理 E.1。我们看到，在非正曲率下不会出现割点。因此，条件密度的表达式中不会出现示性函数，从而保证了更好的光滑性。

• 超球面是紧流形，但对称正定矩阵（n）是非紧的。在紧流形上，向量场 $v(t,x)$ 本身是一致有界的。但在非紧流形上， $\|v(t,x)\|$可能会爆破。因此，在对称正定矩阵（n）上，正则性不太可能逐点成立，但我们仍可期望其范数在期望意义下有界。

• 此外，从高层思路来看，为$v(t,x)$的导数提供上界的步骤保持不变。例如，在引理 D.12中，我们将散度 $v(t,x)$ 的梯度展开为包含 $\mathbb{E}[\|$ 梯度$_x$ 散度 $_x$ 对数映射 $_x(X_1)\| \mid X_t = x]$、 $\mathbb{E}[\|$ 梯度 $_x$ 对数 $p_t(X_1 \mid x)\|^2 \mid X_t = x]$ 等项，仅举几例。要在对称正定矩阵（n）上建立正则性，我们仍大致有相同的展开，涉及同一组项。不同之处在于，我们不再像引理 D.6、D.7、D.8 和 D.9 那样得到一致上界，而是对称正定矩阵（n）上的界将依赖于某个 x_0 的径向距离函数 $r(x) = d(x, x_0)$。

在本节中，我们使用以下记号： $\kappa := \sqrt{-K_{\min}} > 0$，模型函数定义为

$$s_{K_{\min}}(r) := \frac{1}{\kappa}\sinh(\kappa r), \qquad r \geq 0.$$

定义径向收缩映射及其逆映射为

$$\Phi_{t,x_1}(x) := \operatorname{Exp}_{x_1}\big((1-t)\operatorname{Log}_{x_1}(x)\big),$$
$$\Psi_{t,x_1}(x) := \operatorname{Exp}_{x_1}\Big(\frac{1}{1-t}\operatorname{Log}_{x_1}(x)\Big).$$

and 注意，在我们的测地线插值下，$x_t = \Phi_{t,x_1}(x_0)$，以及 $x_0 = \Psi_{t,x_1}(x)$.

我们首先给出密度公式。

引理 E.1 （Hadamard 流形上 $p_t(x_1 \mid x)$ 的密度）。 设(M,g) 为一个完备、单连通、 d维且具有非正曲率的黎曼流形。设 p_0, p_1 分别为先验分布和目标分布，并假设二者相互独立。给定 $X_t = x$ 时， X_1 的条件密度为

$$p_t(x_1 \mid x) = \frac{p_1(x_1)\, p_0\big(\Psi_{t,x_1}(x)\big)\, J_t(x \mid x_1)}{\displaystyle\int_M p_1(z)\, p_0\big(\Psi_{t,z}(x)\big)\, J_t(x \mid z)\, dV_g(z)},$$
$$\textit{where} \quad J_t(x \mid x_1) = (1-t)^{-d}\, \frac{\big|\det(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\big|}{\big|\det(d\operatorname{Exp}_{x_1})_{\operatorname{Log}_{x_1}(x)}\big|}.$$

此外，我们有

$$p_t(x) = \int_M p_1(z)\, p_0\big(\Psi_{t,z}(x)\big)\, J_t(x \mid z)\, dV_g(z).$$

证明。注意，$\Phi_{t,\ x_1}$ 是一个全局微分同胚，而 $\Psi_{t,\ x_1}$ 是其逆映射。对于固定的 x_1 和 x，变量替换公式给出

$$dV_g(\Psi_{t,x_1}(x)) = \big|\det(d\Psi_{t,x_1})_x\big|\, dV_g(x),$$

这意味着

$$p(\Psi_{t,x_1}(x), x_1)\, dV_g(\Psi_{t,x_1}(x))\, dV_g(x_1) = p_0\big(\Psi_{t,x_1}(x)\big)\, p_1(x_1)\, \big|\det(d\Psi_{t,x_1})_x\big|\, dV_g(x)\, dV_g(x_1).$$

我们定义

$$J_t(x \mid x_1) := \big|\det(d\Psi_{t,x_1})_x\big|.$$

等价地，$(X_t,\ X_1)$ 的联合测度可写为

$$p_t(x, x_1)\, dV_g(x)\, dV_g(x_1) = p_0\big(\Psi_{t,x_1}(x)\big)\, p_1(x_1)\, J_t(x \mid x_1)\, dV_g(x)\, dV_g(x_1),$$

对 x_1 积分联合密度可得边缘分布

$$p_t(x) = \int_M p_1(z)\, p_0\big(\Psi_{t,z}(x)\big)\, J_t(x \mid z)\, dV_g(z).$$

因此，根据贝叶斯定理，

$$p_t(x_1 \mid x) = \frac{p_t(x, x_1)}{p_t(x)} = \frac{p_1(x_1)\, p_0\big(\Psi_{t,x_1}(x)\big)\, J_t(x \mid x_1)}{\displaystyle\int_M p_1(z)\, p_0\big(\Psi_{t,z}(x)\big)\, J_t(x \mid z)\, dV_g(z)}.$$

接下来只需计算 $J_t(x \mid x_1)$。回顾 $\Psi_{t,\ x_1}(x)$ = 指数映射 $_{x_1}\Big(\frac{1}{1-t}$ 对数映射 $_{x_1}(x)\Big)$。利用链式法则，

$$(d\Psi_{t,x_1})_x = (d\operatorname{Exp}_{x_1})_v \circ \big((1-t)^{-1} Id\big) \circ (d\operatorname{Log}_{x_1})_x.$$

利用 $(d$ 对数映射 $_{x_1})_x = \big((d$ 指数映射 $_{x_1})_{\operatorname{Log}_{x_1}(x)}\big)^{-1}$，我们有

$$(d\Psi_{t,x_1})_x = (d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)} \circ \big((1-t)^{-1} Id\big) \circ \big((d\operatorname{Exp}_{x_1})_{\operatorname{Log}_{x_1}(x)}\big)^{-1}.$$

两边取行列式，

$$J_t(x \mid x_1) = (d\Psi_{t,x_1})_x = (1-t)^{-d}\, \frac{\big|\det(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\big|}{\big|\det(d\operatorname{Exp}_{x_1})_{\operatorname{Log}_{x_1}(x)}\big|}.$$

□

现在证明命题B.6，这验证了假设4。

证明。 [命题 B 的证明。6] 所需结果可直接由引理 E.4 和引理E.6推出：

$$\begin{aligned}
\mathbb{E}[\|v(t,x)\|^2] &\lesssim \mathbb{E}[d(x_0,x_1)^2] \lesssim d,\\
\mathbb{E}[\|\nabla v(t,x)\|] &\lesssim \frac{d}{1-t} L_R \mathbb{E}[d(x_0,x_1)^2]^{\frac{1}{2}} \mathbb{E}[s_{K_{\min}}(d(x_0,x_1))^6]^{\frac{1}{2}}\\
&\lesssim \frac{d^{2+6\lambda}}{1-t} L_R M_{\lambda_1}^{\frac{1}{2}},\\
\mathbb{E}[|\frac{d}{dt} v(t,x)|] &\lesssim \frac{d}{1-t} L_R \mathbb{E}[d(x_0,x_1)^2 s_{K_{\min}}(d(x_0,x_1))^6]^{\frac{1}{2}} \mathbb{E}[d(x_0,x_1)^2]^{\frac{1}{2}}
\end{aligned}$$

$$\lesssim \frac{d^{2+6\lambda}}{1-t} L_R M_{\lambda_1}^{\frac{1}{2}},$$

$$\mathbb{E}[\| \operatorname{grad}_x \operatorname{div} v(t,x)\|] \lesssim \mathbb{E}[d(x_0,x_1) s_{K_{\min}}(d(x_0,x_1))^6] \frac{d^2}{(1-t)^2} L_R^3 \lesssim \frac{d^{3+12\lambda}}{(1-t)^2} L_R^3,$$

$$\mathbb{E}[|\frac{d}{dt} \operatorname{div} v(t,x)|] \lesssim \mathbb{E}[d(x_0,x_1)^2]^{\frac{1}{2}} \frac{d^2}{(1-t)^2} L_R^3 \mathbb{E}[d(x_1,x_0)^2 s_{K_{\min}}(d(x_0,x_1))^{12}]^{\frac{1}{2}}$$
$$\lesssim \frac{d^{3+12\lambda}}{(1-t)^2} L_R^3 M_{\lambda_1}^{\frac{1}{2}},$$

and

$$\mathbb{E}[\|\nabla v(t,x)\|^2] \lesssim \frac{d^2}{(1-t)^2} L_R^2 \mathbb{E}[d(x_0,x_1)^4]^{\frac{1}{2}} \mathbb{E}[s_{K_{\min}}(d(x_0,x_1))^{12}]^{\frac{1}{2}} \lesssim \frac{d^{3+12\lambda}}{(1-t)^2} L_R^2 M_{\lambda_1}^{\frac{1}{2}},$$

$$\mathbb{E}[|\frac{d}{dt} v(t,x)|^2] \lesssim \frac{d^2}{(1-t)^2} L_R^2 \mathbb{E}[d(x_0,x_1)^4 s_{K_{\min}}(d(x_0,x_1))^{12}]^{\frac{1}{2}} \mathbb{E}[d(x_0,x_1)^4]^{\frac{1}{2}} \lesssim \frac{d^{3+12\lambda}}{(1-t)^2} L_R^2 M_{\lambda_1}^{\frac{1}{2}},$$

$\mathbb{E}[\|$ 梯度 $_x$ 散度 $v(t,x)\|^2] \lesssim \mathbb{E}[d(x_0,x_1)^2 s_{K_{\min}}(d(x_0,x_1))^{12}] \quad \frac{d^4}{(1-t)^4} L_R^6 \lesssim \frac{d^{5+24\lambda}}{(1-t)^4} L_R^6 M_{\lambda_1}.$

□

E.1 命题 5.4 的证明

我们首先证明得分函数正则性结果。在下面的引理中，我们将 $\mathbb{E}[\|$梯度 对数映射 $p_t(X_t)\|^2]$分解为两项之和。

引理 E.2。对于 $X_t \sim p_t$,

$$\mathbb{E}\big[\| \operatorname{grad} \log p_t(X_t)\|^2\big] \le 2\,\mathbb{E}\Big[\big\| \operatorname{grad}_x \log p_0(\Psi_{t,X_1}(X_t))\big\|^2\Big] + 2\,\mathbb{E}\Big[\big\| \operatorname{grad}_x \log J_t(X_t \mid X_1)\big\|^2\Big].$$

证明。 由引理E.1,

$$p_t(x) = \int_M p_1(z)\, p_0\big(\Psi_{t,z}(x)\big)\, J_t(x \mid z)\, dV_g(z).$$

取黎 emannian 梯度关于 x，并在积分号下求导 ,

$$\operatorname{grad} p_t(x) = \int_M p_1(z)\, \operatorname{grad}_x \Big(p_0(\Psi_{t,z}(x))\, J_t(x \mid z)\Big)\, dV_g(z). \tag{17}$$

利用恒等式 grad (ϕ) $= \phi$ grad 对数映射 ϕ，我们有

$$\operatorname{grad}_x \Big(p_0(\Psi_{t,z}(x))\, J_t(x \mid z)\Big) = p_0(\Psi_{t,z}(x))\, J_t(x \mid z)\, \operatorname{grad}_x \log \Big(p_0(\Psi_{t,z}(x))\, J_t(x \mid z)\Big).$$

代入（17）并除以 p_t（x）可得

$$\operatorname{grad} \log p_t(x) = \frac{\operatorname{grad} p_t(x)}{p_t(x)} = \int_M \operatorname{grad}_x \log \Big(p_0(\Psi_{t,z}(x))\, J_t(x \mid z)\Big) \frac{p_1(z)\, p_0(\Psi_{t,z}(x))\, J_t(x \mid z)}{p_t(x)}\, dV_g(z).$$

注意到

$$\frac{p_1(z)\, p_0(\Psi_{t,z}(x))\, J_t(x \mid z)}{p_t(x)} = \frac{p_1(z)\, p_0(\Psi_{t,z}(x))\, J_t(x \mid z)}{\int_M p_1(z)\, p_0\big(\Psi_{t,z}(x)\big)\, J_t(x \mid z)\, dV_g(z)} = p_t(z \mid x).$$

因此我们有

$$\begin{aligned}\operatorname{grad}\,\log p_t(x) = \operatorname{grad}\,\log p_t(x) &= \int_M \operatorname{grad}_x \log\Big(p_0(\Psi_{t,z}(x))\,J_t(x\mid z)\Big)\,p_t(z\mid x)\,dV_g(z)\\ &= \mathbb{E}\Big[\operatorname{grad}_x \log\big(p_0(\Psi_{t,X_1}(x))J_t(x\mid X_1)\big)\mid X_t=x\Big].\end{aligned}$$

接下来，应用詹森不等式，

$$\begin{aligned}\mathbb{E}[\|\operatorname{grad}\,\log p_t(x)\|^2] &= \mathbb{E}\Big[\|\mathbb{E}[\operatorname{grad}_x \log\big(p_0(\Psi_{t,X_1}(x))J_t(x\mid X_1)\big)\mid X_t=x]\|^2\Big]\\ &\le \mathbb{E}\Big[\mathbb{E}[\|\operatorname{grad}_x \log\big(p_0(\Psi_{t,X_1}(x))J_t(x\mid X_1)\big)\|^2\mid X_t=x]\Big]\\ &= \mathbb{E}\Big[\|\operatorname{grad}_x \log\big(p_0(\Psi_{t,X_1}(x))J_t(x\mid X_1)\big)\|^2\Big].\end{aligned}$$

结果由此得出

$$\operatorname{grad}_x \log\big(p_0(\Psi)\,J_t\big) = \operatorname{grad}_x \log p_0(\Psi) + \operatorname{grad}_x \log J_t, \qquad \|a+b\|^2 \le 2\|a\|^2 + 2\|b\|^2.$$

□

现在证明命题5.4。

证明。 [命题 5.4] 由引理 E.2，我们有

$$\mathbb{E}\big[\|\operatorname{grad}\,\log p_t(X_t)\|^2\big] \le 2\,\mathbb{E}\Big[\big\|\operatorname{grad}_x \log p_0(\Psi_{t,X_1}(X_t))\big\|^2\Big] + 2\,\mathbb{E}\Big[\big\|\operatorname{grad}_x \log J_t(X_t\mid X_1)\big\|^2\Big].$$

我们在附录 E.3 （引理E.7 和引理E.8）中证明：

$$\|\operatorname{grad}_x \log J_t(x\mid x_1)\| \le d\,\|(d\operatorname{Log}_{x_1})_x\|\left(\frac{1}{1-t}\,\|\nabla(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\| + \|\nabla(d\operatorname{Exp}_{x_1})_{\operatorname{Log}_{x_1}(x)}\|\right),$$

$$\Big\|\operatorname{grad}_x \log p_0(\Psi_{t,X_1}(X_t))\Big\| \le 2dd(x_0,z)\Big\|(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\Big\|_{\mathrm{op}}\;\frac{1}{1-t}\,\big\|(d\operatorname{Log}_{x_1})_x\big\|_{\mathrm{op}}.$$

我们还在引理 E.10 和引理E.12中证明：

$$\begin{aligned}\|(d\operatorname{Log}_{x_1})_x\| &\le 1,\\ \Big\|(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\Big\|_{\mathrm{op}} &\le s_{K_{\min}}(\|\tfrac{1}{1-t}\operatorname{Log}_{x_1}(x)\|)/\|\tfrac{1}{1-t}\operatorname{Log}_{x_1}(x)\| = s_{K_{\min}}(d(x_0,x_1))/d(x_0,x_1),\\ \|\nabla(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\| &\le \frac{16}{3}s_{K_{\min}}(d(x_0,x_1)/2)^2 L_R s_{K_{\min}}(d(x_0,x_1)).\end{aligned}$$

综上，

$$\begin{aligned}\|\operatorname{grad}_x \log J_t(x\mid x_1)\| &\lesssim \frac{d}{1-t}L_R s_{K_{\min}}(d(x_0,x_1))^3,\\ \Big\|\operatorname{grad}_x \log p_0(\Psi_{t,X_1}(X_t))\Big\| &\lesssim \frac{d}{1-t}s_{K_{\min}}(d(x_0,x_1))d(x_0,z).\end{aligned}$$

根据引理 E.4，我们有

$$\begin{aligned}\mathbb{E}\big[\|\operatorname{grad}\,\log p_t(X_t)\|^2\big] &\lesssim \frac{d^2}{(1-t)^2}L_R^2\mathbb{E}[s_{K_{\min}}(d(x_0,x_1))^6] \lesssim \frac{d^2}{(1-t)^2}L_R^2\mathbb{E}[s_{K_{\min}}(d(x_0,x_1))^6]\\ &\lesssim \frac{d^{2+12\lambda}}{(1-t)^2}L_R^2 M,\end{aligned}$$

其中 $\lambda = \max\{1,\kappa\}$。

□

E.2 Hadamard 流形和 SPD 流形正则性的辅助结果

E.2.1 期望控制

由于 Hadamard流形的非紧致性质以及曲率畸变，要对如下形式的期望进行界定

$$\mathbb{E}[d(X_0,X_1)^2\, s_{K_{\min}}\big(d(X_0,X_1)\big)^a], \quad a\in\mathbb{N},$$

是建立正则性的关键步骤。我们证明，通过选择 X_0 服从黎曼高斯分布，可以对满足特定矩条件（7）的一类数据分布的上述期望进行控制。 X_1

引理 E.3。设$(M、\ g)$为一个完备、单连通的 d 维黎曼流形（*Hadamard*），并假设

$$K_{\min}\ \le\ \sec\ \le\ 0, \qquad where\ K_{\min}<0\ is\ independent\ of\ d.$$

固定一个基点 $z\in M$，并设 $r(x)\colon=d(x、\ z)$。定义

$$p_0(x)=\frac{1}{Z}e^{(-\beta\, r(x)^m)}, \qquad Z:=\int_M e^{-\beta\, r(x)^m}\, dV_g(x), \qquad where \qquad \beta=d, m=2.$$

设 $\lambda\ge 0$ 与 d 相互独立。则对于 $b\in\{0,1,2,3,4\}$,

$$\mathbb{E}_{p_0}\left[r(x)^b e^{\lambda r(x)}\right]\ =\mathcal{O}(d^{2\lambda}(\log d)^b).$$

证明。 我们首先将期望写成切空间上的积分。由于 M 是 Hadamard 流形，$\mathrm{Exp}_z:\ T_zM\ \to\ M$ 是全局 C^∞ 微分同胚。我们在向量空间 T_zM 上赋予内积 g，并将相应的勒贝格测度记为 dv 。对于任意可测的$F\colon\ M\to[0,\infty]$，变量替换公式给出

$$\int_M F(x)\, dV_g(x)=\int_{T_zM} F(\mathrm{Exp}_z(v))\ \left|\det\left(d(\mathrm{Exp}_z)_v\right)\right|\, dv.$$

现在，我们将此公式应用于 $F_1(x)\ =e^{-d\, r(x)^2}$, $F_2(x)\ =r(x)^b e^{\lambda r(x)}e^{-d\, r(x)^2}$。由于 $r(\mathrm{Exp}_z(v))\ =\|v\|$, 我们得到

$$\mathbb{E}_{p_0}[r^b e^{\lambda r}]=\frac{\int_{T_zM}\|v\|^b e^{\lambda\|v\|}e^{-d\|v\|^2}\ \left|\det(d(\mathrm{Exp}_z)_v)\right|\, dv}{\int_{T_zM} e^{-d\|v\|^2}\ \left|\det(d(\mathrm{Exp}_z)_v)\right|\, dv}. \tag{18}$$

回顾比较定理（注意， $d-1$ 来自 det），我们有

$$1\ \le\ \left|\det(d(\mathrm{Exp}_z)_{r\theta})\right|\ \le\ \left(\frac{s_{K_{\min}}(r)}{r}\right)^{d-1}, \qquad \forall r>0. \tag{19}$$

由于 $\frac{s_{K_{\min}}(r)\ r}{}\ =\frac{\sinh(\kappa r)}{\kappa r}\ \le e^{\kappa r}$，我们有

$$\left|\det(d(\mathrm{Exp}_z)_{r\theta})\right|\ \le\ e^{\kappa(d-1)r}. \tag{20}$$

现在我们在 T_zM 使用欧几里得极坐标进行积分： $v=r\theta$ 其中 $r\in[0,\infty)$、 $\theta\in\mathbb{S}^{d-1}$，以及

$$dv=r^{d-1}\, dr\, d\theta.$$

我们在 $R=2$对数映射 d处分段，并对期望进行界定。考虑下面的分解

$$\mathbb{E}_{p_0}[r^b e^{\lambda r}] = \mathbb{E}_{p_0}[r^b e^{\lambda r}\mathbf{1}_{\{r\le R\}}] + \mathbb{E}_{p_0}[r^b e^{\lambda r}\mathbf{1}_{\{r>R\}}].$$

我们首先考虑中心部分。在 $\{r\le R\}$, $r^b e^{\lambda r}\le R^b e^{\lambda R}$，因此

$$\mathbb{E}_{p_0}[r^b e^{\lambda r}\mathbf{1}_{\{r\le R\}}] \le R^b e^{\lambda R} = (2\log d)^b\, d^{2\lambda}. \tag{21}$$

接下来考虑尾部部分。将尾部贡献写作 N_{tail}/Z。利用变量替换公式和上界（20），我们得到

$$\begin{aligned} N_{\text{tail}} &:= \int_{\{x:\, r(x)>R\}} r(x)^b e^{\lambda r(x)} e^{-dr(x)^2}\, dV_g(x) \\ &= \int_{\{v:\, \|v\|>R\}} \|v\|^b e^{\lambda\|v\|} e^{-d\|v\|^2} \left|\det(d(\operatorname{Exp}_z)_v)\right| dv \\ &\le |\mathbb{S}^{d-1}| \int_R^\infty r^{d-1+b} e^{(-dr^2+(\lambda+\kappa(d-1))r)}\, dr. \end{aligned} \tag{22}$$

定义

$$\Psi_d(r) := -dr^2 + (\lambda+\kappa(d-1))r + (d-1+b)\log r.$$

我们证明这个 $\Psi_d(r)$ 可由 $-dr^2$ 项控制。对于 $r\ge R = 2\log d$，有 $(d-1+b)\log r \le \frac{d}{4}r^2$, $\forall d\ge 4$。此外，对于较大的 $r\ge R$和 d， $\kappa(d-1)r\le \frac{d}{4}r^2$，因为这等价于 $r\ge 4\kappa(1-1/d)$，而当 $R\ge 4\kappa$ （即 $d\ge e^{2\kappa}$）时该条件成立。最后，当 $d\ge 8\lambda$ 时，对于所有 $r\ge R$， $\lambda r\le \frac{d}{8}r^2$ 均成立。将这些不等式相加可得，存在某个常数 d_1，使得对所有 $d\ge d_1$及所有 $r\ge R$,

$$(\lambda+\kappa(d-1))r + (d-1+b)\log r \le \frac{d}{2}r^2, \quad \text{hence} \quad \Psi_d(r)\le -\frac{d}{2}r^2.$$

因此，对于 $d\ge d_1$,

$$\begin{aligned} \int_R^\infty e^{\Psi_d(r)}\, dr &\le \int_R^\infty e^{-(d/2)r^2}\, dr = \int_R^\infty \frac{1}{rd}\left(-\frac{d}{dr}e^{-(d/2)r^2}\right) dr \le \frac{1}{Rd}\int_R^\infty \left(-\frac{d}{dr}e^{-(d/2)r^2}\right) dr \\ &= \frac{1}{dR}\exp\Big(-\frac{d}{2}R^2\Big) = \frac{1}{dR}\exp\big(-2d(\log d)^2\big). \end{aligned}$$

代入（22）得

$$N_{\text{tail}} \le |\mathbb{S}^{d-1}|\cdot\frac{1}{dR}\exp\big(-2d(\log d)^2\big). \tag{23}$$

对于分母 Z，注意到 $\left|\det(d(\operatorname{Exp}_z)_v)\right|\ge 1$，因此得到

$$\int_{T_zM} e^{-d\|v\|^2}\left|\det(d(\operatorname{Exp}_z)_v)\right| dv \ge \int_{T_zM} e^{-d\|v\|^2}\, dv = \Big(\frac{\pi}{d}\Big)^{d/2}.$$

因此，利用（23），

$$\mathbb{E}_{p_0}[r^b e^{\lambda r}\mathbf{1}_{\{r>R\}}] = \frac{N_{\text{tail}}}{Z} \le \frac{|\mathbb{S}^{d-1}|}{dR}\exp\big(-2d(\log d)^2\big)\cdot\Big(\frac{d}{\pi}\Big)^{d/2} = \frac{|\mathbb{S}^{d-1}|}{2d\log d}d^{-2d\log d}\cdot\Big(\frac{d}{\pi}\Big)^{d/2} \lesssim 1.$$

综上，将（21）与尾部界结合可得，对于所有 $d\ge d_1$,

$$\mathbb{E}_{p_0}[r^b e^{\lambda r}] \le (2\log d)^b d^{2\lambda} + 1 = \mathcal{O}(d^{2\lambda}(\log d)^b).$$

□

现在我们给出中间步骤。

引理 E.4 (期望的辅助界 $s_{K_{\min}}$)。 设 M 为 *Hadamard* 流形。固定$z \in M$。选择先验分布为 $X_0 \sim p_0$，其中

$$p_0(x) \propto \exp\big(-d\, d(x,z)^2\big).$$

假设条件 *1* 成立。令 $\kappa = \sqrt{-K_{\min}}$和 $\lambda_0 = a$最大值$\{1, \kappa\}$，其中 $0 \le a \le a_0$是某个常数。假设

$$\max\left\{\mathbb{E}[d(X_1,z)^4 e^{\lambda_1 d(X_1,z)}], \mathbb{E}[e^{\lambda_1 d(X_1,z)}]\right\} \le M, \qquad where \qquad \lambda_1 = a_1 \max\{1,\kappa\}.$$

那么，对于 $b \in \{0,1,2,3,4\}$，有

$$\begin{aligned}
\mathbb{E}\Big[d(X_0,X_1)^b\, s_{K_{\min}}\big(d(X_0,X_1)\big)^a\Big] &\lesssim M\mathbb{E}_{X_0\sim p_0}\Big[d(X_0,z)^b e^{\lambda_0 d(X_0,z)}\Big] \lesssim d^{2\lambda_0}(\log d)^b M,\\
\mathbb{E}\Big[d(X_0,z)^b\, s_{K_{\min}}\big(d(X_0,X_1)\big)^a\Big] &\lesssim M\mathbb{E}_{X_0\sim p_0}\Big[d(X_0,z)^b e^{\lambda_0 d(X_0,z)}\Big] \lesssim d^{2\lambda_0}(\log d)^b M.
\end{aligned}$$

证明。 注意到

$$s_{K_{\min}}(d(x_0,x_1))^a = \frac{\sinh^a(\kappa d(x_0,x_1))}{\kappa^a} \le (e^{\max\{1,\kappa\} d(x_0,x_1)})^a.$$

因此

$$d(x_0,x_1)^b\, s_{K_{\min}}\big(d(x_0,x_1)\big)^a \le d(x_0,x_1)^b (e^{\max\{1,\kappa\} d(x_0,x_1)})^a. \tag{24}$$

取期望后，只需对 $\mathbb{E}\big[d(x_0,\ x_1)^b e^{\lambda_0 d(x_0,\ x_1)}\big]$ 进行界定，其中 $\lambda_0 = a$最大值$\{1,\kappa\}$。

由三角不等式可得

$$d(x_0,x_1) \le d(x_0,z) + d(x_1,z).$$

因此,

$$\begin{aligned}
d(x_0,x_1)^b e^{\lambda_0 d(x_0,x_1)} \le& \big(d(x_0,z)+d(x_1,z)\big)^b \exp\big(\lambda_0(d(x_0,z)+d(x_1,z))\big)\\
\lesssim& (d(x_0,z)^b + d(x_1,z)^b)\exp(\lambda_0 d(x_0,z))\exp(\lambda_0 d(x_1,z)).
\end{aligned}$$

取期望并利用 X_0 与 X_1的独立性:

$$\begin{aligned}
\mathbb{E}\Big[d(X_0,X_1)^b e^{\lambda_0 d(X_0,X_1)}\Big] \le& 2\,\mathbb{E}\Big[e^{\lambda_0 d(X_1,z)}\Big]\,\mathbb{E}_{X_0\sim p_0}\Big[d(X_0,z)^b e^{\lambda_0 d(X_0,z)}\Big]\\
&+ 2\,\mathbb{E}\Big[d(X_1,z)^b e^{\lambda_0 d(X_1,z)}\Big]\,\mathbb{E}_{X_0\sim p_0}\Big[e^{\lambda_0 d(X_0,z)}\Big].
\end{aligned} \tag{25}$$

根据假设， $\mathbb{E}\Big[e^{\lambda_0 d\ (X_1,\ z)}\Big]$ 和 $\mathbb{E}\Big[d(X_1,\ z)^b e^{\lambda_0 d\ (X_1,\ z)}\Big]$ 均由 M 控制。因此可得

$$\mathbb{E}\Big[d(X_0,X_1)^b e^{\lambda_0 d(X_0,X_1)}\Big] \lesssim M\mathbb{E}_{X_0\sim p_0}\Big[d(X_0,z)^b e^{\lambda_0 d(X_0,z)}\Big] \lesssim d^{2\lambda_0}(\log d)^b M,$$

其中我们应用了引理 E.3。

□

E.2.2 正则性控制

以下引理总结了建立正则性所需的基本要素，其作用与引理D.6、D.7、D.8和D.9相同。

引理 E.5。 对于 对称正定矩阵（n），其先验分布为 $p_0 \propto$ 指数映射（$-dd$（x_0，z）^{2}），假设假设条件*1*，则有以下界。

$$
\begin{aligned}
\| \operatorname{grad}_x \log J_t(x \mid x_1)\| \lesssim & d \frac{1}{1-t} L_R s_{K_{\min}}(d(x_0,x_1))^3, \\
\left\| \operatorname{grad}_x \log p_0(\Psi_{t,x_1}(x))\right\| \lesssim & d \frac{1}{1-t} s_{K_{\min}}(d(x_0,x_1)) d(x_0,z), \\
|\partial_t \log J_t(x \mid x_1)| \lesssim & \frac{d}{1-t} d(x_0,x_1) s_{K_{\min}}(d(x_0,x_1))^3 L_R, \\
|\partial_t \log p_0(\Psi_{t,x_1}(x))| \lesssim & \frac{d}{1-t} d(x_0,z) s_{K_{\min}}(d(x_0,x_1)), \\
\|\nabla_x^2 \log p_0(\Psi_{t,x_1}(x))\|_{\mathrm{op}} \lesssim & \frac{d}{(1-t)^2} s_{K_{\min}}(d(x_0,x_1))^4 L_R d(x_0,z), \\
\|\nabla_x^2 \log J_t(x \mid x_1)\|_{\mathrm{op}} \lesssim & \frac{d}{(1-t)^2} L_R^3 s_{K_{\min}}(d(x_0,x_1))^6, \\
\left\|\partial_t \operatorname{grad}_x \log p_0(\Psi_{t,x_1}(x))\right\| \lesssim & d(x_0,z) \frac{d}{(1-t)^2} d(x_0,x_1) L_R s_{K_{\min}}(d(x_0,x_1))^3, \\
\|\partial_t \operatorname{grad}_x \log J_t(x \mid x_1)\| \lesssim & \frac{d}{(1-t)^2} L_R^3 d(x_1,x_0) s_{K_{\min}}(d(x_0,x_1))^6, \\
\|\nabla_x \operatorname{Log}_x(x_1)\|_{\mathrm{op}}^2 \lesssim & d(x_0,x_1)^2, \\
|\operatorname{div}_x \operatorname{Log}_x(x_1)|^2 \lesssim & d^2 d(x_0,x_1)^2, \\
\| \operatorname{grad} \operatorname{div}_x \operatorname{Log}_x(x_1)\| \lesssim & d d(x_0,x_1)^{\frac{3}{2}}.
\end{aligned}
$$

这里我们强调，对于固定的 x_1，对数映射 $_x$（x_1）被视为一个向量场，因此 ∇对数映射 $_x$（x_1）是向量场 对数映射$_x$（x_1）的协变导数。因此，

$$
\begin{aligned}
&\mathbb{E}[\| \text{梯度}_x \text{对数映射 } p_t(x_1 \mid x)\|^2 \mid X_t = x] \lesssim \frac{d^2}{(1-t)} {}_2 L_R^2 \mathbb{E}[s_{K_{\min}}(d(x_0, x_1))^6], \\
&\mathbb{E}[\| \text{梯度}_x \text{对数映射 } p_t(x_1 \mid x)\|^4 \mid X_t = x] \lesssim \frac{d^4}{(1-t)} {}_4 L_R^4 \mathbb{E}[s_{K_{\min}}(d(x_0, x_1))^{12}], \\
&\mathbb{E}[|\partial_t \text{对数映射 } p_t(x_1 \mid x)|^2 \mid X_t = x] \lesssim \frac{d^2}{(1-t)} {}_2 L_R^2 \mathbb{E}[d(x_0, x_1)^2 s_{K_{\min}}(d(x_0, x_1))^6], \\
&\mathbb{E}[|\partial_t \text{对数映射 } p_t(x_1 \mid x)|^4 \mid X_t = x] \lesssim \frac{d^4}{(1-t)} {}_4 L_R^4 \mathbb{E}[d(x_0, x_1)^4 s_{K_{\min}}(d(x_0, x_1))^{12}], \\
&\mathbb{E}[\|\nabla_x^2 \text{对数映射 } p_t(x_1 \mid x)\|_{\mathrm{op}}^2 \mid X_t = x] \lesssim \frac{d^2}{(1-t)} {}_4 L_R^6 \mathbb{E}[s_{K_{\min}}(d(x_0, x_1))^{12}] \\
&\qquad + \frac{d^4}{(1-t)^4} L_R^4 \mathbb{E}[s_{K_{\min}}(d(x_0,x_1))^6]^2, \\
&\mathbb{E}[\|\nabla_x^2 \text{对数映射 } p_t(x_1 \mid x)\|_{\mathrm{op}}^4 \mid X_t = x] \lesssim \frac{d^4}{(1-t)} {}_8 L_R^{12} \mathbb{E}[s_{K_{\min}}(d(x_0, x_1))^{24}] \\
&\qquad + \frac{d^8}{(1-t)^8} L_R^8 \mathbb{E}[s_{K_{\min}}(d(x_0,x_1))^{12}]^2,
\end{aligned}
$$

$$\mathbb{E}[\|\partial_t \operatorname{grad}_x \log p_t(x_1 \mid x)\|^2 \mid X_t = x] \lesssim \frac{d^2}{(1-t)^4} L_R^6 \mathbb{E}[d(x_1,x_0)^2 s_{K_{\min}}(d(x_0,x_1))^{12}] + \frac{d^4}{(1-t)^4} L_R^4 \mathbb{E}[s_{K_{\min}}(d(x_0,x_1))^6]\mathbb{E}[d(x_0,x_1)^2 s_{K_{\min}}(d(x_0,x_1))^6].$$

证明。在整个证明过程中，我们用 r（x）表示径向距离函数 d（x，z），其中 z是 $p_0 \propto$ 指数映射 $(-\beta d$（x_0，z）^{2}的中心。我们在附录E.3（引理E.7 和引理E.8）中证明了以下结论：

$$\|\operatorname{grad}_x \log J_t(x \mid x_1)\| \le d\,\|(d\operatorname{Log}_{x_1})_x\| \left(\frac{1}{1-t} \|\nabla(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\| + \|\nabla(d\operatorname{Exp}_{x_1})_{\operatorname{Log}_{x_1}(x)}\| \right),$$

$$\left\|\operatorname{grad}_x \log p_0(\Psi_{t,X_1}(X_t))\right\| \le 2d d_g(x_0,z) \left\|(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\right\|_{\mathrm{op}} \frac{1}{1-t} \left\|(d\operatorname{Log}_{x_1})_x\right\|_{\mathrm{op}}.$$

我们还在以下内容中证明：引理 E.10 和引理E.12：

$$\begin{aligned}
\|(d\operatorname{Log}_{x_1})_x\| &\le 1,\\
\left\|\left(\nabla(d\operatorname{Log}_x)\right)_y\right\| &\le \left\|\left(\nabla(d\operatorname{Exp}_x)\right)_{\operatorname{Log}_x(y)}\right\|,\\
\left\|(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\right\|_{\mathrm{op}} &\le s_{K_{\min}}(\|\tfrac{1}{1-t}\operatorname{Log}_{x_1}(x)\|)/\|\tfrac{1}{1-t}\operatorname{Log}_{x_1}(x)\| \lesssim s_{K_{\min}}(d(x_0,x_1)),\\
\|\nabla(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\| &\le \frac{16}{3} s_{K_{\min}}(d(x_0,x_1)/2)^2 L_R s_{K_{\min}}(d(x_0,x_1)) \lesssim L_R s_{K_{\min}}(d(x_0,x_1))^3,\\
\|\nabla^2(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\| &\lesssim L_R^3 s_{K_{\min}}(d(x_0,x_1))^5.
\end{aligned}$$

我们按如下方式估计各项。对于梯度和时间导数，我们有

$$\begin{aligned}
\|\operatorname{grad}_x \log J_t(x \mid x_1)\| &\lesssim d\,\frac{1}{1-t} L_R s_{K_{\min}}(d(x_0,x_1))^3,\\
\left\|\operatorname{grad}_x \log p_0(\Psi_{t,X_1}(X_t))\right\| &\lesssim d \frac{1}{1-t} s_{K_{\min}}(d(x_0,x_1)) d(x_0,z),
\end{aligned}$$

and

$$\begin{aligned}
|\partial_t \log J_t(x \mid x_1)| &\le \frac{d}{1-t} + \frac{d}{(1-t)^2} \|\operatorname{Log}_{x_1}(x)\| \, \|\nabla(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\|\\
&\lesssim \frac{d}{1-t} d(x_0,x_1) s_{K_{\min}}(d(x_0,x_1))^3 L_R,\\
|\partial_t \log p_0(\Psi_{t,x_1}(x))| &\le 2\beta r \left\|(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\right\|_{\mathrm{op}} \frac{1}{(1-t)^2} \|\operatorname{Log}_{x_1}(x)\|\\
&\lesssim \frac{d}{1-t} d(x_0,z) s_{K_{\min}}(d(x_0,x_1)).
\end{aligned}$$

对于二阶协变导数，

$$\begin{aligned}
&\|\nabla_x^2 \log p_0(\Psi_{t,x_1}(x))\|_{\mathrm{op}}\\
\le &\left(\left\|(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\right\|_{\mathrm{op}} \frac{1}{1-t} \left\|(d\operatorname{Log}_{x_1})_x\right\|_{\mathrm{op}}\right)^2 \left(2\beta + 2\beta r \|\nabla^2 r\|_{\mathrm{op}}\Big|_{\Psi_{t,x_1}(x)}\right)\\
&+ \left\|\nabla(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\right\| \left(\frac{1}{1-t}\left\|(d\operatorname{Log}_{x_1})_x\right\|_{\mathrm{op}}\right)^2 2\beta r
\end{aligned}$$

$$
\begin{aligned}
&+\left\|(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\right\|_{\mathrm{op}}\frac{1}{1-t}\left\|\nabla(d\operatorname{Log}_{x_1})_x\right\|2\beta r\\
\lesssim&\frac{1}{(1-t)^2}s_{K_{\min}}(d(x_0,x_1))^2(dd(x_0,z)\kappa\coth(\kappa d(x_0,z)))\\
&+\frac{d}{(1-t)^2}s_{K_{\min}}(d(x_0,x_1))^3L_Rd(x_0,z)+\frac{d}{1-t}d(x_0,z)s_{K_{\min}}(d(x_0,x_1))^4L_R\\
\lesssim&\frac{d}{(1-t)^2}s_{K_{\min}}(d(x_0,x_1))^4L_Rd(x_0,z),
\end{aligned}
$$

其中注意到 $x\coth(x)$ 为 x 阶。并且我们有

$$
\begin{aligned}
&\|\nabla_x^2\log J_t(x\mid x_1)\|_{\mathrm{op}}\\
\le&d\left\|(d\operatorname{Log}_{x_1})_x\right\|^2\Bigg[\frac{1}{(1-t)^2}\Big(\|\nabla(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\|^2+\|\nabla^2(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\|\Big)\\
&+\Big(\|\nabla(d\operatorname{Exp}_{x_1})_{\operatorname{Log}_{x_1}(x)}\|^2+\|\nabla^2(d\operatorname{Exp}_{x_1})_{\operatorname{Log}_{x_1}(x)}\|\Big)\Bigg]\\
&+d\left\|\nabla(d\operatorname{Log}_{x_1})_x\right\|\left(\frac{1}{1-t}\|\nabla(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\|+\|\nabla(d\operatorname{Exp}_{x_1})_{\operatorname{Log}_{x_1}(x)}\|\right)\\
\lesssim&\frac{d}{(1-t)^2}(L_R^2s_{K_{\min}}(d(x_0,x_1))^6+L_R^3s_{K_{\min}}(d(x_0,x_1))^5)+\frac{d}{1-t}L_R^2s_{K_{\min}}(d(x_0,x_1))^6\\
\lesssim&\frac{d}{(1-t)^2}L_R^3s_{K_{\min}}(d(x_0,x_1))^6.
\end{aligned}
$$

对于梯度的时间导数,

$$
\begin{aligned}
&\left\|\partial_t\operatorname{grad}_x\log p_0(\Psi_{t,x_1}(x))\right\|\\
\le&\Big(\left\|(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\right\|_{\mathrm{op}}\frac{1}{1-t}\left\|(d\operatorname{Log}_{x_1})_x\right\|_{\mathrm{op}}\Big)\times\Big(2\beta+2\beta r\|\nabla^2r\|_{\mathrm{op}}\Big|_{\Psi_{t,x_1}(x)}\Big)\\
&\times\left\|(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\right\|_{\mathrm{op}}\frac{1}{(1-t)^2}\left\|\operatorname{Log}_{x_1}(x)\right\|\\
&+\Bigg[\frac{1}{(1-t)^2}\left\|(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\right\|_{\mathrm{op}}\left\|(d\operatorname{Log}_{x_1})_x\right\|_{\mathrm{op}}\\
&+\frac{1}{(1-t)^3}\left\|\operatorname{Log}_{x_1}(x)\right\|\left\|\nabla(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\right\|\left\|(d\operatorname{Log}_{x_1})_x\right\|_{\mathrm{op}}\Bigg]2\beta r\\
\lesssim&\Big(d+dd(x_0,z)\Big)s_{K_{\min}}(d(x_0,x_1))^2\frac{1}{(1-t)^2}\\
&+dd(x_0,z)\Big(\frac{1}{(1-t)^2}s_{K_{\min}}(d(x_0,x_1))+\frac{1}{(1-t)^2}d(x_0,x_1)L_Rs_{K_{\min}}(d(x_0,x_1))^3\Big)\\
\lesssim&d(x_0,z)\frac{d}{(1-t)^2}d(x_0,x_1)L_Rs_{K_{\min}}(d(x_0,x_1))^3.
\end{aligned}
$$

并且我们有

$\|\partial_t$ 梯度$_x$ 对数映射 $J_t(x\mid x_1)\|$

$$
\begin{aligned}
\leq & \|(d\operatorname{Log}_{x_1})_x\|\left[\frac{d}{(1-t)^2}\|\nabla(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\|\right. \\
& \left.+\frac{d}{(1-t)^3}\|\operatorname{Log}_{x_1}(x)\|\left(\|\nabla(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\|^2+\|\nabla^2(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\|\right)\right] \\
\lesssim & \frac{d}{(1-t)^2}L_R s_{K_{\min}}(d(x_0,x_1))^3+\frac{d}{(1-t)^2}d(x_1,x_0)\left(L_R^2 s_{K_{\min}}(d(x_0,x_1))^6+L_R^3 s_{K_{\min}}(d(x_0,x_1))^5\right) \\
\lesssim & \frac{d}{(1-t)^2}L_R^3 d(x_1,x_0)s_{K_{\min}}(d(x_0,x_1))^6.
\end{aligned}
$$

Log 上的三个不等式由附录E.5得出:

$$
\begin{aligned}
\|\nabla_x \operatorname{Log}_x(y)\|_{\mathrm{op}}^2 &\leq 2+2\,d(x,y)^2\Big(\frac{s'_{K_{\min}}(d(x,y))}{s_{K_{\min}}(d(x,y))}\Big)^2, \\
|\operatorname{div}_x \operatorname{Log}_x(y)|^2 &\leq 2+2(d-1)^2\,d(x,y)^2\Big(\frac{s'_{K_{\min}}(d(x,y))}{s_{K_{\min}}(d(x,y))}\Big)^2, \\
\|\operatorname{grad}\operatorname{div}_x \operatorname{Log}_x(y)\| &\leq \frac{\sqrt{2}d}{2}(2\big(1+d(x,y)\,\frac{s'_{K_{\min}}(d(x,y))}{s_{K_{\min}}(d(x,y))}\big))^{\frac{3}{2}}.
\end{aligned}
$$

现在我们证明关于 $\log p_t$ 的导数的不等式。我们指出，由于引理 E.4，比较 $\log\ J_t$ 和 $\log\ p_0$ 的导数的逐点上界时，由于 $d(x_0,z)$ 项不会比 $d(x_0,x_1)$ 项提高期望的阶数，我们可以将 $d(x_0,z)$ 视为 $d(x_0,x_1)$，并看到 $\log\ J_t$ 的导数支配 $\log\ p_0$ 的相同阶导数。因此，在计算 $\log p_t=\log p_0(\Psi_{t,x_1}(x))+\log\ J_t(x|x_1)+\text{const}$ 的导数时，只需考虑 $\log J_t(x|x_1)$ 项的导数。

前四个不等式通过应用柯西-施瓦茨不等式即可直接得到，并注意到它们被 $\log J_t$ 项所支配。根据引理 D.8

$$
\begin{aligned}
\|\nabla_x^2\log p_t(x_1\mid x)\|_{\mathrm{op}}^2 \lesssim & \|\nabla^2\log J_t(x\mid X_1)\|_{\mathrm{op}}^2+\mathbb{E}\big[\|\nabla^2\log J_t(x\mid X_1)\|_{\mathrm{op}}^2\mid X_t=x\big] \\
& +\mathbb{E}\big[\|\operatorname{grad}_x\log J_t(x\mid X_1)\|^2\mid X_t=x\big]\mathbb{E}\big[\|\operatorname{grad}\log p_t(x_1\mid x)\|^2\mid X_t=x\big], \\
\|\nabla_x^2\log p_t(x_1\mid x)\|_{\mathrm{op}}^4 \lesssim & \|\nabla^2\log J_t(x\mid X_1)\|_{\mathrm{op}}^4 \\
& +\mathbb{E}\big[\|\operatorname{grad}_x\log J_t(x\mid X_1)\|^4\mid X_t=x\big]\mathbb{E}\big[\|\operatorname{grad}\log p_t(x_1\mid x)\|^4\mid X_t=x\big],
\end{aligned}
$$

并代入相应的表达式（使用柯西-施瓦茨不等式），我们得到 $\log p_t$的海森矩阵的上界。根据引理 D.9,

$$
\begin{aligned}
\|\partial_t\operatorname{grad}_x\log p_t(x_1\mid x)\| \leq & \|\partial_t\operatorname{grad}_x\log J_t(x\mid x_1)\|+\mathbb{E}\big[\|\partial_t\operatorname{grad}_x\log J_t(x\mid X_1)\|\mid X_t=x\big] \\
& +\mathbb{E}[\|\operatorname{grad}_x\log J_t(x\mid X_1)\|\,|\partial_t\log p_t(X_1\mid x)|\mid X_t=x],
\end{aligned}
$$

我们代入相　　应的表达式（使用柯西-施瓦茨不等式），即可得到最后一个不等式。　□

我们在下面的引理中总结所需的界，该引理与引理 D.10、D.11、D.12 和 D.13类似。

引理 E.6。在对称正定矩阵 (n)上，先验分布为 $p_0\propto$指数映射$(-dd(x_0,\ z)^2)$，在假设 *1* 的条件下，我们有

$$
\mathbb{E}[\|v(t,x)\|^2]\lesssim\mathbb{E}[d(x_0,x_1)^2],
$$

$$
\begin{aligned}
\mathbb{E}[\|\nabla v(t,x)\|] \lesssim & \frac{d}{1-t} L_R \mathbb{E}[d(x_0,x_1)^2]^{\frac{1}{2}} \mathbb{E}[s_{K_{\min}}(d(x_0,x_1))^6]^{\frac{1}{2}}, \\
\mathbb{E}[|\frac{d}{dt} v(t,x)|] \lesssim & \frac{d}{1-t} L_R \mathbb{E}[d(x_0,x_1)^2 s_{K_{\min}}(d(x_0,x_1))^6]^{\frac{1}{2}} \mathbb{E}[d(x_0,x_1)^2]^{\frac{1}{2}}, \\
\mathbb{E}[\|\operatorname{grad}_x \operatorname{div} v(t,x)\|] \lesssim & \mathbb{E}[d(x_0,x_1) s_{K_{\min}}(d(x_0,x_1))^6] \frac{d^2}{(1-t)^2} L_R^3, \\
\mathbb{E}[|\frac{d}{dt} \operatorname{div} v(t,x)|] \lesssim & \mathbb{E}[d(x_0,x_1)^2]^{\frac{1}{2}} \frac{d^2}{(1-t)^2} L_R^3 \mathbb{E}[d(x_1,x_0)^2 s_{K_{\min}}(d(x_0,x_1))^{12}]^{\frac{1}{2}}.
\end{aligned}
$$

此外，

$$
\begin{aligned}
\mathbb{E}[\|\nabla v(t,x)\|^2] \lesssim & \frac{d^2}{(1-t)^2} L_R^2 \mathbb{E}[d(x_0,x_1)^4]^{\frac{1}{2}} \mathbb{E}[s_{K_{\min}}(d(x_0,x_1))^{12}]^{\frac{1}{2}}, \\
\mathbb{E}[|\frac{d}{dt} v(t,x)|^2] \lesssim & \frac{d^2}{(1-t)^2} L_R^2 \mathbb{E}[d(x_0,x_1)^4 s_{K_{\min}}(d(x_0,x_1))^{12}]^{\frac{1}{2}} \mathbb{E}[d(x_0,x_1)^4]^{\frac{1}{2}}, \\
\mathbb{E}[\|\text{梯度}_x \text{散度}\ v(t,x)\|^2] \lesssim & \mathbb{E}[d(x_0,x_1)^2 s_{K_{\min}}(d(x_0,x_1))^{12}] \quad \frac{d^4}{(1-t)^4} L_R^6.
\end{aligned}
$$

证明。 在整个证明过程中，我们将使用引理 E.5中的界。我们首先控制向量场正则性。这与引理D.10和引理D.11类似。我们有

$$
\begin{aligned}
\mathbb{E}[\|\nabla v(t,x)\|] \leq & \frac{1}{1-t} \mathbb{E}\big[\|\operatorname{Log}_x(X_1)\| \, \|\operatorname{grad}_x \log p_t(X_1 \mid x)\| \mid X_t = x\big] \\
& + \frac{1}{1-t} \mathbb{E}[\|\nabla \operatorname{Log}_x(x_1)\| \mid X_t = x] \\
\lesssim & \frac{d}{1-t} L_R \mathbb{E}[d(x_0,x_1)^2]^{\frac{1}{2}} \mathbb{E}[s_{K_{\min}}(d(x_0,x_1))^6]^{\frac{1}{2}},
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbb{E}[|\frac{d}{dt} v(t,x)|] = & \frac{1}{(1-t)^2} \mathbb{E}[\operatorname{Log}_x(x_1) \mid X_t = x] + \frac{1}{1-t} \mathbb{E}[\operatorname{Log}_x(x_1) \, \partial_t \log p_t(x_1 \mid x) \mid X_t = x] \\
\lesssim & \frac{d}{1-t} L_R \mathbb{E}[d(x_0,x_1)^2 s_{K_{\min}}(d(x_0,x_1))^6]^{\frac{1}{2}} \mathbb{E}[d(x_0,x_1)^2]^{\frac{1}{2}}.
\end{aligned}
$$

现在我们控制散度正则性。与引理 D.12类似，我们有

$$
\|\operatorname{grad}_x \operatorname{div} v(t,x)\| \leq \frac{1}{1-t} \Big(T_1 + T_2 + T_3 + T_4 \Big),
$$

其中

$$
\begin{aligned}
\mathbb{E}[T_1] = & \mathbb{E}[\|\operatorname{grad}_x \operatorname{div}_x \operatorname{Log}_x(X_1)\|] \\
\lesssim & d \mathbb{E}[d(x_0,x_1)^{\frac{3}{2}}], \\
\mathbb{E}[T_2] = & \mathbb{E}[|\operatorname{div}_x \operatorname{Log}_x(X_1)| \, \|\operatorname{grad}_x \log p_t(X_1 \mid x)\|] \\
\lesssim & \mathbb{E}[|\operatorname{div}_x \operatorname{Log}_x(X_1)|^2]^{\frac{1}{2}} \mathbb{E}[\|\operatorname{grad}_x \log p_t(X_1 \mid x)\|^2]^{\frac{1}{2}} \\
\lesssim & \frac{d^2}{1-t} L_R \mathbb{E}[d(x_0,x_1)^2]^{\frac{1}{2}} \mathbb{E}[s_{K_{\min}}(d(x_0,x_1))^6]^{\frac{1}{2}}, \\
\mathbb{E}[T_3] = & \mathbb{E}[\|\operatorname{grad}_x \langle \operatorname{grad}_x \log p_t(X_1 \mid x), \operatorname{Log}_x(X_1) \rangle\|]
\end{aligned}
$$

$$
\begin{aligned}
&\le \mathbb{E}[\|\nabla \operatorname{Log}_x(X_1)\| \, \|\operatorname{grad}_x \log p_t(X_1 \mid x)\|] \\
&\qquad + \mathbb{E}[\|\nabla^2 \log p_t(X_1 \mid x)\| \, \|\operatorname{Log}_x(X_1)\|] \\
&\lesssim \Big(\mathbb{E}[d(x,x_1)^2]\frac{d^2}{(1-t)^2}L_R^2\mathbb{E}[s_{K_{\min}}(d(x_0,x_1))^6]\Big)^{\frac{1}{2}} \\
&\quad + \Big(\frac{d^2}{(1-t)^4}L_R^6\mathbb{E}[s_{K_{\min}}(d(x_0,x_1))^{12}]\mathbb{E}[d(x,x_1)^2]\Big)^{\frac{1}{2}} \\
&\lesssim \frac{d}{1-t}L_R^3\mathbb{E}[s_{K_{\min}}(d(x_0,x_1))^{12}]^{\frac{1}{2}}\mathbb{E}[d(x_0,x_1)^2]^{\frac{1}{2}}, \\
\mathbb{E}[T_4] =& \mathbb{E}[|\langle \operatorname{grad}_x \log p_t(X_1 \mid x), \operatorname{Log}_x(X_1)\rangle| \, \|\operatorname{grad}_x \log p_t(X_1 \mid x)\|] \\
&\le \mathbb{E}[\|\operatorname{Log}_x(X_1)\| \, \|\operatorname{grad}_x \log p_t(X_1 \mid x)\|^2] \\
&\lesssim \mathbb{E}[d(x,x_1)s_{K_{\min}}(d(x_0,x_1))^6]\frac{d^2}{(1-t)^2}L_R^2 \\
&\lesssim \mathbb{E}[d(x_0,x_1)s_{K_{\min}}(d(x_0,x_1))^6]\frac{d^2}{1-t}L_R^2.
\end{aligned}
$$

与引理D.13类似

$$
\begin{aligned}
\frac{d}{dt}\operatorname{div}\, v(t,x) =& \frac{1}{(1-t)^2}\mathbb{E}\big[\operatorname{div}_x \operatorname{Log}_x(X_1) + \langle \operatorname{grad}_x \log p_t(X_1 \mid x), \operatorname{Log}_x(X_1)\rangle \mid X_t = x\big] \\
&+ \frac{1}{1-t}\mathbb{E}[\operatorname{div}_x \operatorname{Log}_x(X_1)\, \partial_t \log p_t(X_1 \mid x) \mid X_t = x] \\
&+ \frac{1}{1-t}\mathbb{E}[\partial_t \langle \operatorname{grad}_x \log p_t(X_1 \mid x), \operatorname{Log}_x(X_1)\rangle \mid X_t = x] \\
&+ \frac{1}{1-t}\mathbb{E}[\langle \operatorname{grad}_x \log p_t(X_1 \mid x), \operatorname{Log}_x(X_1)\rangle\, \partial_t \log p_t(X_1 \mid x) \mid X_t = x].
\end{aligned}
$$

我们界定

$$
\begin{aligned}
\mathbb{E}[T_1] =& \frac{1}{(1-t)^2}\mathbb{E}\big[\operatorname{div}_x \operatorname{Log}_x(X_1) + \langle \operatorname{grad}_x \log p_t(X_1 \mid x), \operatorname{Log}_x(X_1)\rangle\big] \\
&\lesssim \frac{d}{(1-t)^2}\mathbb{E}[d(x,x_1)] + \frac{1}{(1-t)^2}\mathbb{E}[d(x,x_1)^2]^{\frac{1}{2}}\mathbb{E}[\frac{d^2}{(1-t)^2}L_R^2 s_{K_{\min}}(d(x_0,x_1))^6]^{\frac{1}{2}} \\
&\lesssim \frac{d}{1-t}\mathbb{E}[d(x_0,x_1)] + \frac{d}{(1-t)^2}L_R\mathbb{E}[d(x_0,x_1)^2]^{\frac{1}{2}}\mathbb{E}[s_{K_{\min}}(d(x_0,x_1))^6]]^{\frac{1}{2}}, \\
\mathbb{E}[T_2] =& \frac{1}{1-t}\mathbb{E}[\text{散度}_x \text{对数映射}_x(X_1)\ \partial_t \text{对数}\ p_t(X_1 \mid x)] \\
&\lesssim \frac{1}{1-t}d\mathbb{E}[d(x,x_1)^2]^{\frac{1}{2}}\frac{d}{1-t}L_R\mathbb{E}[d(x_0,x_1)^2 s_{K_{\min}}(d(x_0,x_1))^6]^{\frac{1}{2}} \\
&\lesssim \frac{d^2}{1-t}L_R\mathbb{E}[d(x_0,x_1)^2]^{\frac{1}{2}}\mathbb{E}[d(x_0,x_1)^2 s_{K_{\min}}(d(x_0,x_1))^6]^{\frac{1}{2}}, \\
\mathbb{E}[T_3] =& \frac{1}{1-t}\mathbb{E}[\partial_t \langle \text{梯度}_x \text{对数}\ p_t(X_1 \mid x),\ \text{对数映射}_x(X_1)\rangle] \\
&\lesssim \mathbb{E}[d(x_0,x_1)^2]^{\frac{1}{2}}\frac{d}{(1-t)^2}L_R^3\mathbb{E}[d(x_1,x_0)^2 s_{K_{\min}}(d(x_0,x_1))^{12}]^{\frac{1}{2}}, \\
\mathbb{E}[T_4] =& \frac{1}{1-t}\mathbb{E}[\langle \text{梯度}_x \text{对数映射}\ p_t(X_1 \mid x), \text{对数映射}_x(X_1)\rangle\, \partial_t \text{对数映射}\ p_t(X_1 \mid x)] \\
&\lesssim \frac{1}{1-t}(\frac{d^3}{(1-t)^3}L_R^3\mathbb{E}[s_{K_{\min}}(d(x_0,x_1))^9])^{\frac{1}{3}}\mathbb{E}[d(x,x_1)^3]^{\frac{1}{3}}(\frac{d^3}{(1-t)^3}L_R^3\mathbb{E}[d(x_0,x_1)^3 s_{K_{\min}}(d(x_0,x_1))^9])^{\frac{1}{3}}
\end{aligned}
$$

$$\lesssim \frac{d^2}{(1-t)^2} L_R^2 \mathbb{E}[s_{K_{\min}}(d(x_0,x_1))^9]^{\frac{1}{3}} \mathbb{E}[d(x_0,x_1)^3]^{\frac{1}{3}} \mathbb{E}[d(x_0,x_1)^3 s_{K_{\min}}(d(x_0,x_1))^9]^{\frac{1}{3}}.$$

高阶界可通过完全相同的过程得到。

□

E.3 导数的辅助界

回顾一下，我们使用以下记号

$$\Psi_{t,x_1}(x) = \operatorname{Exp}_{x_1}\Big(\tfrac{1}{1-t}\operatorname{Log}_{x_1}(x)\Big).$$

引理 E.7（控制对数映射的导数 p_0）。设 M 为 *Hadamard* 流形。固定 $\beta > 0$、$m = 2$ 和 $z \in M$，并定义

$$p_0(x) \propto \exp\big(-\beta\, d(x,z)^m\big).$$

固定 $t \in [0,1)$、$x_1 \in M$ 和 $x \in M$。定义 $r := d(\Psi_{t,\,x_1}(x),\ z)$，使得 r 表示 x_0 与 z 之间的径向距离。则以下界成立：

$$\begin{aligned}
\Big\|\operatorname{grad}_x \log p_0 \Psi_{t,x_1}(x)\Big\| \le& 2\beta r \Big\|(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\Big\|_{\mathrm{op}} \ \frac{1}{1-t}\ \big\|(d\operatorname{Log}_{x_1})_x\big\|_{\mathrm{op}},\\
|\partial_t \log p_0(\Psi_{t,x_1}(x))| \le& 2\beta r \Big\|(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\Big\|_{\mathrm{op}} \ \frac{1}{(1-t)^2}\ \big\|\operatorname{Log}_{x_1}(x)\big\|,\\
\|\nabla_x^2 \log p_0(\Psi_{t,x_1}(x))\|_{\mathrm{op}} \le& \Big(\Big\|(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\Big\|_{\mathrm{op}} \ \frac{1}{1-t}\ \big\|(d\operatorname{Log}_{x_1})_x\big\|_{\mathrm{op}}\Big)^2\\
&\Big(2\beta + 2\beta r \|\nabla^2 r\|_{\mathrm{op}}\Big|_{\Psi_{t,x_1}(x)}\Big)\\
&+ \Big\|\nabla (d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\Big\| \Big(\frac{1}{1-t}\big\|(d\operatorname{Log}_{x_1})_x\big\|_{\mathrm{op}}\Big)^2 2\beta r\\
&+ \Big\|(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\Big\|_{\mathrm{op}} \ \frac{1}{1-t}\ \big\|\nabla (d\operatorname{Log}_{x_1})_x\big\| 2\beta r,\\
\Big\|\partial_t \operatorname{grad}_x \log p_0(\Psi_{t,x_1}(x))\Big\| \le& \Big(\Big\|(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\Big\|_{\mathrm{op}} \ \frac{1}{1-t}\ \big\|(d\operatorname{Log}_{x_1})_x\big\|_{\mathrm{op}}\Big)\\
&\times \Big(2\beta + 2\beta r \|\nabla^2 r\|_{\mathrm{op}}\Big|_{\Psi_{t,x_1}(x)}\Big)\\
&\times \Big\|(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\Big\|_{\mathrm{op}} \ \frac{1}{(1-t)^2}\ \big\|\operatorname{Log}_{x_1}(x)\big\|\\
&+ \Bigg[\frac{1}{(1-t)^2}\Big\|(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\Big\|_{\mathrm{op}} \ \big\|(d\operatorname{Log}_{x_1})_x\big\|_{\mathrm{op}}\\
&+ \frac{1}{(1-t)^3}\big\|\operatorname{Log}_{x_1}(x)\big\| \Big\|\nabla (d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\Big\| \ \big\|(d\operatorname{Log}_{x_1})_x\big\|_{\mathrm{op}}\Bigg] 2\beta r.
\end{aligned}$$

其中

$$\|\nabla^2 r\|_{\mathrm{op}}\Big|_{\Psi_{t,x_1}(x)} \le \frac{s'_{K_{\min}}(r)}{s_{K_{\min}}(r)} = \sqrt{-K_{\min}}\coth(r\sqrt{-K_{\min}}).$$

并注意到 $r \mapsto r\cosh(r)$ 在 $r = 0$ 处没有奇点。

证明。 我们首先推导 $\|$梯度 对数映射 $p_0(\Psi_{t,x_1}(x))\|$ 和 $\|\nabla^2$ 对数映射 $p_0(\Psi_{t,x_1}(x))\|_{\mathrm{op}}$ 的界。由于 $\log\ p_0(y) = -\beta\, d(y,z)^m +$ 为常数，且 $\|$ 梯度 $d(\cdot\,,z)\| = 1$，因此我们有

$$\|\operatorname{grad}\log p_0(\Psi_{t,x_1}(x))\| = 2\beta r.$$

此外，请注意

$$\nabla^2(d(y,z)^m) = 2d(y,z)\nabla^2 d(y,z) + 2\operatorname{grad}\, d(y,z)\otimes\operatorname{grad}\, d(y,z).$$

上述恒等式，结合 $\|$梯度 $d(y,z)\otimes$ 梯度 $d(y,z)\|_{\mathrm{op}} = 1$，可得

$$\|\nabla^2\log p_0(\Psi_{t,x_1}(x))\|_{\mathrm{op}} \le 2\beta + 2\beta\, r\,\|\nabla^2 r\|_{\mathrm{op}}\Big|_{\Psi_{t,x_1}(x)}.$$

接下来我们推导Ψ_{t,x_1}的第一个（空间）导数。利用链式法则，

$$(d\Psi_{t,x_1})_x[u] = (d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\Big[\tfrac{1}{1-t}\,(d\operatorname{Log}_{x_1})_x[u]\Big], \qquad \forall\, u\in T_xM.$$

因此

$$\|(d\Psi_{t,x_1})_x\|_{\mathrm{op}} \le \Big\|(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\Big\|_{\mathrm{op}}\ \tfrac{1}{1-t}\,\big\|(d\operatorname{Log}_{x_1})_x\big\|_{\mathrm{op}}. \tag{26}$$

对于任意 $u\in T_xM$，链式法则给出恒等式

$$\big\langle \operatorname{grad}_x\log p_0(\Psi_{t,x_1}(x)),\, u\big\rangle = \big\langle \operatorname{grad}\,\log p_0(\Psi_{t,x_1}(x)),\, (d\Psi_{t,x_1})_x[u]\big\rangle.$$

因此，根据柯西-施瓦茨不等式并对 $\|u\| = 1$ 取上确界，可得

$$\begin{aligned}\big|\langle\operatorname{grad}_x\log p_0(\Psi_{t,x_1}(x)),u\rangle\big| &\le \|\operatorname{grad}\,\log p_0(\Psi_{t,x_1}(x))\|\,\|(d\Psi_{t,x_1})_x[u]\|\\ &\le \|\operatorname{grad}\,\log p_0(\Psi_{t,x_1}(x))\|\,\|(d\Psi_{t,x_1})_x\|_{\mathrm{op}}.\end{aligned}$$

代入 $\|$梯度对数 $p_0(\Psi_{t,x_1}(x))\| = 2\beta r$ 并利用（26），我们得到

$$\begin{aligned}&\Big\|\operatorname{grad}_x\log p_0\Big(\operatorname{Exp}_{x_1}\big(\tfrac{1}{1-t}\operatorname{Log}_{x_1}(x)\big)\Big)\Big\|\\ \le\ & 2\beta r\,\Big\|(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\Big\|_{\mathrm{op}}\ \frac{1}{1-t}\,\big\|(d\operatorname{Log}_{x_1})_x\big\|_{\mathrm{op}}.\end{aligned}$$

现在我们处理一阶时间导数。回顾 $\Psi_{t,\ x_1}(x)$ = 指数映射 $_{x_1}\Big(\frac{1}{1-t}$对数映射 $_{x_1}(x)\Big)$。根据链式法则，

$$\partial_t\log p_0(\Psi_{t,x_1}(x)) = \big\langle \operatorname{grad}\log p_0(\Psi_{t,x_1}(x)),\, \partial_t\Psi_{t,x_1}(x)\big\rangle,$$

其中 $\partial_t\Psi_{t,\ x_1}(x)$ = $(d$ 指数映射 $_{x_1})_{\frac{1}{1-t}\text{对数映射}_{x_1}(x)}\Big[\frac{1}{1-t()^2}$对数映射 $_{x_1}(x)\Big]$. 根据柯西-施瓦茨不等式，

$$\begin{aligned}|\partial_t\log p_0(\Psi_{t,x_1}(x))| &\le \|\operatorname{grad}\,\log p_0(\Psi_{t,x_1}(x))\|\,\|\partial_t\Psi_{t,x_1}(x)\|\\ &\le \|\operatorname{grad}\,\log p_0(\Psi_{t,x_1}(x))\|\,\Big\|(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\Big\|_{\mathrm{op}}\ \tfrac{1}{(1-t)^2}\,\|\operatorname{Log}_{x_1}(x)\|\\ &\le 2\beta r\Big\|(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\Big\|_{\mathrm{op}}\ \frac{1}{(1-t)^2}\,\|\operatorname{Log}_{x_1}(x)\|,\end{aligned}$$

其中回忆一下 $\|$梯度 log $p_0(\Psi_{t,\ x_1}(x))\| = 2\beta r$。

现在我们计算海森矩阵。对于任意 $u,\ w \in T_xM$，利用链式法则，我们得到 n

$$\begin{aligned}\nabla_x^2 \log p_0(\Psi_{t,x_1}(x))[u,w] =& \nabla^2 \log p_0(\Psi_{t,x_1}(x))\big[(\Psi_{t,x_1})_x[u], (\Psi_{t,x_1})_x[w]\big] \\ &+ \big\langle \operatorname{grad} \log p_0(\Psi_{t,x_1}(x)),\, \nabla_u\big((\Psi_{t,x_1})_x[w]\big)\big\rangle.\end{aligned}$$

根据柯西-施瓦茨不等式，并对 $\|u\| = \|w\| = 1$ 取上确界，我们得到

$$\begin{aligned}&\|\nabla_x^2 \log p_0(\Psi_{t,x_1}(x))\|_{\mathrm{op}} = \sup_{\|u\|=\|w\|=1} \big|\nabla_x^2 \log p_0(\Psi_{t,x_1}(x))[u,w]\big| \\ \le& \sup_{\|u\|=\|w\|=1} \|\nabla^2 \log p_0(\Psi_{t,x_1}(x))\|_{\mathrm{op}}\, \|(\Psi_{t,x_1})_x[u]\|\, \|(\Psi_{t,x_1})_x[w]\| \\ &+ \|\operatorname{grad} \log p_0(\Psi_{t,x_1}(x))\|\, \big\|\nabla_u\big((\Psi_{t,x_1})_x[w]\big)\big\| \\ \le& \|\nabla^2 \log p_0(\Psi_{t,x_1}(x))\|_{\mathrm{op}}\, \|(\Psi_{t,x_1})_x\|_{\mathrm{op}}^2 + \|\operatorname{grad} \log p_0(\Psi_{t,x_1}(x))\|\, \|\nabla(\Psi_{t,x_1})_x\|.\end{aligned}$$

接下来需要界定 $\|\nabla(\Psi_{t,\ x_1})_x\|$。回顾

$$(\Psi_{t,x_1})_x[w] = (d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\Big[\tfrac{1}{1-t}(d\operatorname{Log}_{x_1})_x[w]\Big].$$

回顾莱布尼茨法则，其中 $v = \frac{1}{1-t}\operatorname{Log}_{x_1}(x)$ 是位于 $T_{x_1}M$ 处的一个依赖于 x 的切向量。

$$\begin{aligned}\nabla_u\big((\Psi_{t,x_1})_x[w]\big) &= \nabla_u\Big((d\operatorname{Exp}_{x_1})_v\Big[\tfrac{1}{1-t}(d\operatorname{Log}_{x_1})_x[w]\Big]\Big) \\ &= \Big(\nabla_{\nabla_u v}(d\operatorname{Exp}_{x_1})\Big)_v\Big[\tfrac{1}{1-t}(d\operatorname{Log}_{x_1})_x[w]\Big] \ + \ (d\operatorname{Exp}_{x_1})_v\Big[\tfrac{1}{1-t}\nabla_u\big((d\operatorname{Log}_{x_1})_x[w]\big)\Big].\end{aligned}$$

其中注意到 v 依赖于 x，因此利用链式法则，

$$\begin{aligned}\nabla_u((d\operatorname{Exp}_{x_1})_v) &= \nabla_u((d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}) \\ &= \big(\nabla(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\big)(\nabla_u \tfrac{1}{1-t}\operatorname{Log}_{x_1}(x)) \\ &= \Big(\nabla_{\nabla_u \frac{1}{1-t}\operatorname{Log}_{x_1}(x)}(d\operatorname{Exp}_{x_1})\Big)_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}.\end{aligned}$$

注意，对于 $v = \frac{1}{1-t}\operatorname{Log}_{x_1}(x)$，其一阶导数是一个方向导数。我们有

$$\nabla_u v = \tfrac{1}{1-t}(d\operatorname{Log}_{x_1})_x[u], \qquad \text{and} \qquad \nabla_u\big((d\operatorname{Log}_{x_1})_x[w]\big) = \nabla(d\operatorname{Log}_{x_1})_x[u,w].$$

取范数并应用算子范数的定义可得

$$\begin{aligned}\big\|\nabla_u\big((\Psi_{t,x_1})_x[w]\big)\big\| &\le \Big\|\nabla(d\operatorname{Exp}_{x_1})_v\Big\|\, \|\nabla_u v\|\, \Big\|\tfrac{1}{1-t}(d\operatorname{Log}_{x_1})_x[w]\Big\| \\ &\quad + \ \Big\|(d\operatorname{Exp}_{x_1})_v\Big\|_{\mathrm{op}}\, \tfrac{1}{1-t}\, \big\|\nabla(d\operatorname{Log}_{x_1})_x[u,w]\big\| \\ &= \Big\|\nabla(d\operatorname{Exp}_{x_1})_v\Big\|\, \Big\|\tfrac{1}{1-t}(d\operatorname{Log}_{x_1})_x[u]\Big\|\, \Big\|\tfrac{1}{1-t}(d\operatorname{Log}_{x_1})_x[w]\Big\| \\ &\quad + \Big\|(d\operatorname{Exp}_{x_1})_v\Big\|_{\mathrm{op}}\, \tfrac{1}{1-t}\, \big\|\nabla(d\operatorname{Log}_{x_1})_x[u,w]\big\|.\end{aligned}$$

对 $\|u\| = \|w\| = 1$ 取上确界，我们得到

$$\big\|\nabla\big((d\Psi_{t,x_1})_x\big)\big\| \le \Big\|\nabla(d\operatorname{Exp}_{x_1})_v\Big\|\, \Big(\tfrac{1}{1-t}\big\|(d\operatorname{Log}_{x_1})_x\big\|_{\mathrm{op}}\Big)^2 + \Big\|(d\operatorname{Exp}_{x_1})_v\Big\|_{\mathrm{op}}\, \tfrac{1}{1-t}\, \big\|\nabla(d\operatorname{Log}_{x_1})_x\big\|.$$

最后利用以下事实

$$\begin{aligned}&\|\operatorname{grad}\log p_0(\Psi_{t,x_1}(x))\| = 2\beta r,\\ &\|\nabla^2\log p_0(\Psi_{t,x_1}(x))\|_{\mathrm{op}} \le 2\beta + 2\beta r\|\nabla^2 r\|_{\mathrm{op}}\big|_{\Psi_{t,x_1}(x)},\end{aligned}$$

并利用（26），我们得到

$$\begin{aligned}&\|\nabla_x^2\log p_0(\Psi_{t,x_1}(x))\|_{\mathrm{op}}\\ \le&\|\nabla^2\log p_0(\Psi_{t,x_1}(x))\|_{\mathrm{op}}\,\|(\Psi_{t,x_1})_x\|_{\mathrm{op}}^2 + \|\operatorname{grad}\log p_0(\Psi_{t,x_1}(x))\|\,\|\nabla(\Psi_{t,x_1})_x\|\\ \le&\Big(\Big\|(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\Big\|_{\mathrm{op}}\,\frac{1}{1-t}\,\big\|(d\operatorname{Log}_{x_1})_x\big\|_{\mathrm{op}}\Big)^2\Big(2\beta+2\beta r\|\nabla^2 r\|_{\mathrm{op}}\Big|_{\Psi_{t,x_1}(x)}\Big)\\ &+\Big\|\nabla(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\Big\|\Big(\frac{1}{1-t}\big\|(d\operatorname{Log}_{x_1})_x\big\|_{\mathrm{op}}\Big)^2 2\beta r\\ &+\Big\|(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\Big\|_{\mathrm{op}}\,\frac{1}{1-t}\,\big\|\nabla(d\operatorname{Log}_{x_1})_x\big\|\,2\beta r.\end{aligned}$$

剩　下需要考虑混合导数项。对于任意 $u\in T_xM$，回忆我们有

$$\big\langle \operatorname{grad}_x\log p_0(\Psi_{t,x_1}(x)),u\big\rangle = \big\langle \operatorname{grad}\log p_0(\Psi_{t,x_1}(x)),(\Psi_{t,x_1})_x[u]\big\rangle.$$

对两边关于 t 求导（固定 x_1,x,u）：

$$\begin{aligned}\big\langle\partial_t\operatorname{grad}_x\log p_0(\Psi_{t,x_1}(x)),u\big\rangle =&\Big\langle\nabla^2\log p_0(\Psi_{t,x_1}(x))\big[\partial_t\Psi_{t,x_1}(x)\big],\,(\Psi_{t,x_1})_x[u]\Big\rangle\\ &+\big\langle\operatorname{grad}\log p_0(\Psi_{t,x_1}(x)),\,\partial_t\big((\Psi_{t,x_1})_x[u]\big)\big\rangle.\end{aligned}$$

根据柯西-施瓦茨不等式，

$$\begin{aligned}\big|\langle\partial_t\operatorname{grad}_x\log p_0(\Psi_{t,x_1}(x)),u\rangle\big| \le&\|\nabla^2\log p_0(\Psi_{t,x_1}(x))\|_{\mathrm{op}}\,\|\partial_t\Psi_{t,x_1}(x)\|\,\|(\Psi_{t,x_1})_x\|_{\mathrm{op}}\,\|u\|\\ &+\|\operatorname{grad}\log p_0(\Psi_{t,x_1}(x))\|\,\Big\|\partial_t\big((\Psi_{t,x_1})_x[u]\big)\Big\|.\end{aligned}$$

根据前面的步骤，我们已经得到 $\|\partial_t\Psi_{t,x_1}(x)\|$和 $\|(\Psi_{t,x_1})_x\|_{\mathrm{op}}$。

接下来需要对 $\|\partial_t((\Psi_{t,\,x_1})_x[u])\|$ 进行界定。回顾一下，对于任意 $u\in T_xM$，

$$(\Psi_{t,x_1})_x[u] = (d\operatorname{Exp}_{x_1})_v\Big[\tfrac{1}{1-t}(d\operatorname{Log}_{x_1})_x[u]\Big],\qquad v=\tfrac{1}{1-t}\operatorname{Log}_{x_1}(x).$$

固定 x_1，x，u，并对 t 求导。由于 x 是固定的，张量 $(d$对数映射$_{x_1})_x[u]$ 不依赖于 t；只有标量因子 $\frac{1}{1-t}$ 和点 v 依赖于 t。根据莱布尼茨法则（将依赖于 t 的线性映射作用于依赖于 t 的向量时所用的乘积法则），

$$\partial_t\big((\Psi_{t,x_1})_x[u]\big) = \partial_t\Big((d\operatorname{Exp}_{x_1})_v\Big)\Big[\tfrac{1}{1-t}(d\operatorname{Log}_{x_1})_x[u]\Big] + (d\operatorname{Exp}_{x_1})_v\Big[\partial_t\Big(\tfrac{1}{1-t}(d\operatorname{Log}_{x_1})_x[u]\Big)\Big].$$

对于第一项，$(d$指数映射$_{x_1})_v$ 仅通过 $v(t)$ 依赖于 t，因此根据链式法则，

$$\partial_t\Big((d\operatorname{Exp}_{x_1})_v\Big) = \Big(\nabla(d\operatorname{Exp}_{x_1})_v\Big)(\partial_t v) = \Big(\nabla_{\partial_t v}(d\operatorname{Exp}_{x_1})\Big)_v.$$

对于第二项，由于 $(d$对数映射$_{x_1})_x[u]$ 与 t无关，

$$\partial_t\Big(\tfrac{1}{1-t}(d\operatorname{Log}_{x_1})_x[u]\Big) = \tfrac{1}{(1-t)^2}(d\operatorname{Log}_{x_1})_x[u].$$

结合这些恒等式可得公式

$$\partial_t\big((\Psi_{t,x_1})_x[u]\big)=\Big(\nabla_{\partial_t v}(d\operatorname{Exp}_{x_1})\Big)_v\Big[\tfrac{1}{1-t}(d\operatorname{Log}_{x_1})_x[u]\Big]+(d\operatorname{Exp}_{x_1})_v\Big[\tfrac{1}{(1-t)^2}(d\operatorname{Log}_{x_1})_x[u]\Big].$$

注意到 $\partial_t v=\frac{1}{(1-t)^2}$ 对数映射$x_1(x)$，取范数可得（对于 $\|u\|=1$）

$$\begin{aligned}
&\Big\|\partial_t\big((\Psi_{t,x_1})_x[u]\big)\Big\|\\
\le&\Big\|\nabla(d\operatorname{Exp}_{x_1})_v\Big\|\ \|\partial_t v\|\ \Big\|\tfrac{1}{1-t}(d\operatorname{Log}_{x_1})_x[u]\Big\|\\
&\quad+\Big\|(d\operatorname{Exp}_{x_1})_v\Big\|_{\mathrm{op}}\ \tfrac{1}{(1-t)^2}\ \big\|(d\operatorname{Log}_{x_1})_x\big\|_{\mathrm{op}}\ \|u\|\\
=&\Big\|\nabla(d\operatorname{Exp}_{x_1})_v\Big\|\ \tfrac{1}{(1-t)^2}\|\operatorname{Log}_{x_1}(x)\|\ \tfrac{1}{1-t}\big\|(d\operatorname{Log}_{x_1})_x\big\|_{\mathrm{op}}\\
&\quad+\Big\|(d\operatorname{Exp}_{x_1})_v\Big\|_{\mathrm{op}}\ \tfrac{1}{(1-t)^2}\ \big\|(d\operatorname{Log}_{x_1})_x\big\|_{\mathrm{op}}\\
=&\tfrac{1}{(1-t)^3}\|\operatorname{Log}_{x_1}(x)\|\ \Big\|\nabla(d\operatorname{Exp}_{x_1})_v\Big\|\ \big\|(d\operatorname{Log}_{x_1})_x\big\|_{\mathrm{op}}\\
&\quad+\tfrac{1}{(1-t)^2}\Big\|(d\operatorname{Exp}_{x_1})_v\Big\|_{\mathrm{op}}\ \big\|(d\operatorname{Log}_{x_1})_x\big\|_{\mathrm{op}}.
\end{aligned}$$

代入 ‖梯度 对数映射 $p_0(\Psi_{t,x_1}(x))\|$ 和 $\|\nabla^2$ 对数映射 $p_0(\Psi_{t,x_1}(x))\|_{\mathrm{op}}$ 的表达式，得到

$$\begin{aligned}
&\Big\|\partial_t\operatorname{grad}_x\log p_0\Big(\operatorname{Exp}_{x_1}\big(\tfrac{1}{1-t}\operatorname{Log}_{x_1}(x)\big)\Big)\Big\|\\
\le&\Big(\Big\|(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\Big\|_{\mathrm{op}}\ \frac{1}{1-t}\ \big\|(d\operatorname{Log}_{x_1})_x\big\|_{\mathrm{op}}\Big)\\
&\times\Big(2\beta+2\beta r\|\nabla^2 r\|_{\mathrm{op}}\Big|_{\Psi_{t,x_1}(x)}\Big)\\
&\times\Big\|(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\Big\|_{\mathrm{op}}\ \frac{1}{(1-t)^2}\ \big\|\operatorname{Log}_{x_1}(x)\big\|\\
&+\Bigg[\frac{1}{(1-t)^2}\Big\|(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\Big\|_{\mathrm{op}}\ \big\|(d\operatorname{Log}_{x_1})_x\big\|_{\mathrm{op}}\\
&+\frac{1}{(1-t)^3}\big\|\operatorname{Log}_{x_1}(x)\big\|\ \Big\|\nabla(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\Big\|\ \big\|(d\operatorname{Log}_{x_1})_x\big\|_{\mathrm{op}}\Bigg]2\beta r.
\end{aligned}$$

最后，我们指出，当 $K_{\min}<0$ 时，Hessian 比较定理给出

$$\|\nabla^2 r\|_{\mathrm{op}}\Big|_{\Psi_{t,x_1}(x)}\le\frac{s'_{K_{\min}}(r)}{s_{K_{\min}}(r)}=\sqrt{-K_{\min}}\coth(r\sqrt{-K_{\min}}).$$

□

引理 E.8（控制 J_t 项的导数）。 假设 M 是一个 *Hadamard* 流形。固定 $x_1\in M$、$x\in M$ 和 $t\in$（0，1）。回忆

$$\begin{aligned}
J_t(x\mid x_1)&=(1-t)^{-d}\,\frac{\big|\det(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\big|}{\big|\det(d\operatorname{Exp}_{x_1})_{\operatorname{Log}_{x_1}(x)}\big|},\\
\log J_t(x\mid x_1)&=-d\log(1-t)+\log\det(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}-\log\det(d\operatorname{Exp}_{x_1})_{\operatorname{Log}_{x_1}(x)}.
\end{aligned}$$

于是，以下逐点界成立。

$$
\begin{aligned}
\|\operatorname{grad}_x \log J_t(x \mid x_1)\| \le & d\,\|(d\operatorname{Log}_{x_1})_x\|\left(\frac{1}{1-t}\,\|\nabla(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\| + \|\nabla(d\operatorname{Exp}_{x_1})_{\operatorname{Log}_{x_1}(x)}\|\right),\\
\|\nabla_x^2 \log J_t(x \mid x_1)\|_{\mathrm{op}} \le & d\,\|(d\operatorname{Log}_{x_1})_x\|^2\left[\frac{1}{(1-t)^2}\left(\|\nabla(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\|^2\right.\right.\\
& \left.+\|\nabla^2(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\|\right)\\
& \left.+\left(\|\nabla(d\operatorname{Exp}_{x_1})_{\operatorname{Log}_{x_1}(x)}\|^2+\|\nabla^2(d\operatorname{Exp}_{x_1})_{\operatorname{Log}_{x_1}(x)}\|\right)\right]\\
& + d\,\|\nabla(d\operatorname{Log}_{x_1})_x\|\left(\frac{1}{1-t}\,\|\nabla(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\| + \|\nabla(d\operatorname{Exp}_{x_1})_{\operatorname{Log}_{x_1}(x)}\|\right),\\
|\partial_t \log J_t(x \mid x_1)| \le & \frac{d}{1-t} + \frac{d}{(1-t)^2}\,\|\operatorname{Log}_{x_1}(x)\|\,\|\nabla(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\|,\\
\|\partial_t \operatorname{grad}_x \log J_t(x \mid x_1)\| \le & \|(d\operatorname{Log}_{x_1})_x\|\left[\frac{d}{(1-t)^2}\|\nabla(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\|\right.\\
& + \frac{d}{(1-t)^3}\,\|\operatorname{Log}_{x_1}(x)\|\\
& \left.\left(\|\nabla(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\|^2+\|\nabla^2(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\|\right)\right].
\end{aligned}
$$

证明。 我们反复使用引理 E.16，其中 $A(\cdot)$ 取为 $\xi \mapsto (d\text{指数映射}_{x_1})\xi$，并通过对数映射使用链式法则$_{x_1}$。

我们首先估计梯度。由

$$\log J_t(x \mid x_1) = -d\log(1-t) + \log\det(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)} - \log\det(d\operatorname{Exp}_{x_1})_{\operatorname{Log}_{x_1}(x)},$$

常数 $-d\log(1-t)$ 的 x-梯度为零，因此

$$\operatorname{grad}_x \log J_t(x \mid x_1) = \operatorname{grad}_x \log\det(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)} - \operatorname{grad}_x \log\det(d\operatorname{Exp}_{x_1})_{\operatorname{Log}_{x_1}(x)}.$$

考虑第一项。对映射 $x \mapsto \frac{1}{1-t}$对数映射 $_{x_1}(x)$ 求导会产生因子 $\frac{1}{1-t}$(d对数映射 $_{x_1})_x$。因此，

$$
\begin{aligned}
&\left\|\operatorname{grad}_x \log\det(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\right\|\\
\le &\frac{1}{1-t}\,\|(d\operatorname{Log}_{x_1})_x\|\,\left\|\operatorname{grad}_\xi \log\det(d\operatorname{Exp}_{x_1})_\xi\right\|\Big|_{\xi=\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}.
\end{aligned}
$$

现在应用 Lemma E.16（具体而言，式（33）中的梯度界）于 $A(\xi) = (d\text{指数映射}_{x_1})\xi$ 到 g et

$$\left\|\operatorname{grad}_\xi \log\det(d\operatorname{Exp}_{x_1})_\xi\right\| \le d\,\|(d\operatorname{Exp}_{x_1})_\xi^{-1}\|\,\|\nabla(d\operatorname{Exp}_{x_1})_\xi\|.$$

综合以上，我们得到

$$\left\|\operatorname{grad}_x \log\det(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\right\|$$

$$\leq \frac{d}{1-t}\,\|(d\operatorname{Log}_{x_1})_x\|\,\|(d\operatorname{Exp}_{x_1})^{-1}_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\|\,\|\nabla(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\|.$$

同样的论证（不含因子 $\frac{1}{1-t}$）可得

$$\left\|\operatorname{grad}_x \log\det(d\operatorname{Exp}_{x_1})_{\operatorname{Log}_{x_1}(x)}\right\| \leq d\,\|(d\operatorname{Log}_{x_1})_x\|\,\|(d\operatorname{Exp}_{x_1})^{-1}_{\operatorname{Log}_{x_1}(x)}\|\,\|\nabla(d\operatorname{Exp}_{x_1})_{\operatorname{Log}_{x_1}(x)}\|.$$

最后，应用三角不等式，可得

$$\begin{aligned}\|\operatorname{grad}_x \log J_t(x\mid x_1)\| \leq d\,\|(d\operatorname{Log}_{x_1})_x\|\Bigg(&\frac{1}{1-t}\,\|(d\operatorname{Exp}_{x_1})^{-1}_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\|\,\|\nabla(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\|\\ &+\|(d\operatorname{Exp}_{x_1})^{-1}_{\operatorname{Log}_{x_1}(x)}\|\,\|\nabla(d\operatorname{Exp}_{x_1})_{\operatorname{Log}_{x_1}(x)}\|\Bigg).\end{aligned}$$

现在估计海森矩阵。记 $f(x) := \log\det(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\text{对数映射}_{x_1}(x)}$。取二阶导数（使用乘积法则），可得

$$\begin{aligned}\|\nabla^2_x f(x)\|_{\mathrm{op}} \leq&\left\|\nabla^2_\xi \log\det(d\operatorname{Exp}_{x_1})_\xi\right\|_{\mathrm{op}}\Bigg|_{\xi=\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\cdot\left\|\frac{1}{1-t}(d\operatorname{Log}_{x_1})_x\right\|^2\\ &+\left\|\operatorname{grad}_\xi \log\det(d\operatorname{Exp}_{x_1})_\xi\right\|\Bigg|_{\xi=\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\cdot\left\|\nabla\left(\frac{1}{1-t}(d\operatorname{Log}_{x_1})_x\right)\right\|.\end{aligned}$$

因此有（这里 ∇^2_ξ 表示对 ξ 求导）

$$\begin{aligned}&\|\nabla^2_x \log J_t(x\mid x_1)\|_{\mathrm{op}}\\ \leq&\|(d\operatorname{Log}_{x_1})_x\|^2\left(\frac{1}{(1-t)^2}\left\|\nabla^2_\xi \log\det(d\operatorname{Exp}_{x_1})_\xi\right\|_{\mathrm{op}}\Bigg|_{\xi=\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}+\left\|\nabla^2_\xi \log\det(d\operatorname{Exp}_{x_1})_\xi\right\|_{\mathrm{op}}\Bigg|_{\xi=\operatorname{Log}_{x_1}(x)}\right)\\ &+\|\nabla(d\operatorname{Log}_{x_1})_x\|\left(\frac{1}{1-t}\left\|\operatorname{grad}_\xi \log\det(d\operatorname{Exp}_{x_1})_\xi\right\|\Bigg|_{\xi=\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}+\left\|\operatorname{grad}_\xi \log\det(d\operatorname{Exp}_{x_1})_\xi\right\|\Bigg|_{\xi=\operatorname{Log}_{x_1}(x)}\right)\\ \leq&d\,\|(d\operatorname{Log}_{x_1})_x\|^2\Bigg[\frac{1}{(1-t)^2}\Big(\|(d\operatorname{Exp}_{x_1})^{-1}_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\|^2\,\|\nabla(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\|^2\\ &+\|(d\operatorname{Exp}_{x_1})^{-1}_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\|\,\|\nabla^2(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\|\Big)\\ &+\Big(\|(d\operatorname{Exp}_{x_1})^{-1}_{\operatorname{Log}_{x_1}(x)}\|^2\,\|\nabla(d\operatorname{Exp}_{x_1})_{\operatorname{Log}_{x_1}(x)}\|^2+\|(d\operatorname{Exp}_{x_1})^{-1}_{\operatorname{Log}_{x_1}(x)}\|\,\|\nabla^2(d\operatorname{Exp}_{x_1})_{\operatorname{Log}_{x_1}(x)}\|\Big)\Bigg]\\ &+d\,\|\nabla(d\operatorname{Log}_{x_1})_x\|\left(\frac{1}{1-t}\,\|(d\operatorname{Exp}_{x_1})^{-1}_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\|\,\|\nabla(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\|\right.\\ &\left.+\|(d\operatorname{Exp}_{x_1})^{-1}_{\operatorname{Log}_{x_1}(x)}\|\,\|\nabla(d\operatorname{Exp}_{x_1})_{\operatorname{Log}_{x_1}(x)}\|\right),\end{aligned}$$

其中，第一个不等式使用了三角不等式，第二个不等式应用了引理 E.16。

对于时间导数，由于 x 是固定的，只有 $\frac{1}{1-t}$ 取决于 t，

$$\partial_t \log J_t(x \mid x_1) = \frac{d}{1-t} + \partial_t\Big(\log\det(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\Big).$$

根据链式法则，

$$\partial_t\Big(\log\det(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\Big) = D\log\det(d\operatorname{Exp}_{x_1})_\xi\Big|_{\xi=\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\Big[\partial_t\Big(\frac{1}{1-t}\operatorname{Log}_{x_1}(x)\Big)\Big].$$

但 $\partial_t\big(\frac{1}{1-t}\operatorname{Log}_{x_1}(x)\big) = \frac{1}{1-t()^2}\operatorname{Log}_{x_1}(x)$，因此

$$\Big|\partial_t\Big(\log\det(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\Big)\Big| \le \frac{1}{(1-t)^2}\,\|\operatorname{Log}_{x_1}(x)\|\,\Big\|\operatorname{grad}_\xi\log\det(d\operatorname{Exp}_{x_1})_\xi\Big\|\Big|_{\xi=\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}.$$

应用引理 E.16 （梯度界（33））于 $\xi = \frac{1}{1-t}\operatorname{Log}_{x_1}(x)$，得到

$$|\partial_t \log J_t(x \mid x_1)| \le \frac{d}{1-t} + \frac{d}{(1-t)^2}\,\|\operatorname{Log}_{x_1}(x)\|\,\|(d\operatorname{Exp}_{x_1})^{-1}_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\|\,\|\nabla(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\|.$$

对于混合导数，回顾

$$\operatorname{grad}_x \log J_t(x \mid x_1) = \operatorname{grad}_x \log\det(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)} - \operatorname{grad}_x \log\det(d\operatorname{Exp}_{x_1})_{\operatorname{Log}_{x_1}(x)}.$$

只有第一项依赖于 t。对第一项求导：有两部分，一部分来自对前因子 $\frac{1}{1-t}$ 求导，另一部分来自对 $\operatorname{grad}_\xi \log\det(d\operatorname{Exp}_{x_1})_\xi$ at $\xi = \frac{1}{1-t}\operatorname{Log}_{x_1}(x)$ 求导。这得到

$$\begin{aligned}
&\|\partial_t \operatorname{grad}_x \log\det(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\| \\
\le&\|(d\operatorname{Log}_{x_1})_x\|\Bigg[\frac{1}{(1-t)^2}\,\Big\|\operatorname{grad}_\xi\log\det(d\operatorname{Exp}_{x_1})_\xi\Big\|\Big|_{\xi=\frac{1}{1-t}\operatorname{Log}_{x_1}(x)} \\
&+\frac{1}{1-t}\,\Big\|\partial_t\Big(\operatorname{grad}_\xi\log\det(d\operatorname{Exp}_{x_1})_\xi\Big)\Big\|\Big|_{\xi=\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\Bigg].
\end{aligned}$$

接下来，注意到

$$\partial_t(\operatorname{grad}_\xi \log\det(d\operatorname{Exp}_{x_1})_\xi) = \nabla^2_\xi \log\det(d\operatorname{Exp}_{x_1})_\xi\,[\partial_t\xi]$$

其中 $\partial_t\xi = \frac{1}{(1-t)^2}$ 对数映射 $\operatorname{Log}_{x_1}(x)$。因此

$$\Big\|\partial_t\Big(\operatorname{grad}_\xi\log\det(d\operatorname{Exp}_{x_1})_\xi\Big)\Big\| \le \frac{1}{(1-t)^2}\,\|\operatorname{Log}_{x_1}(x)\|\,\Big\|\nabla^2_\xi\log\det(d\operatorname{Exp}_{x_1})_\xi\Big\|_{\mathrm{op}}.$$

根据引理 E.16，我们得到

$$\begin{aligned}
&\|\partial_t \operatorname{grad}_x \log J_t(x \mid x_1)\| \\
\le&\|(d\operatorname{Log}_{x_1})_x\|\Bigg[\frac{1}{(1-t)^2}\,d\,\|(d\operatorname{Exp}_{x_1})^{-1}_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\|\,\|\nabla(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\|
\end{aligned}$$

$$+\frac{1}{(1-t)^3}\|\operatorname{Log}_{x_1}(x)\|\, d\Big(\|(d\operatorname{Exp}_{x_1})^{-1}_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\|^2\,\|\nabla(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\|^2$$
$$+\|(d\operatorname{Exp}_{x_1})^{-1}_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\|\,\|\nabla^2(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\|\Big)\Bigg].$$

最后，根据 Lezcano-Casado (2020，定理 3.12) 及 $r = \frac{\|\text{对数映射}_{x_1}(x)\|}{1-t}$，并注意到在 SPD($n$) 上，截面曲率的上界为零，我们得到

$$\|(d\operatorname{Exp}_{x_1})^{-1}_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\| \le 1.$$

因此我们得到

$$\|\operatorname{grad}_x \log J_t(x\mid x_1)\| \le d\,\|(d\operatorname{Log}_{x_1})_x\|\left(\frac{1}{1-t}\|\nabla(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\| + \|\nabla(d\operatorname{Exp}_{x_1})_{\operatorname{Log}_{x_1}(x)}\|\right),$$

$\|\nabla^2_x \text{对数映射}\ J_t(x\mid x_1)\|_{\text{op}} \le d\,\|(d\text{对数映射}_{x_1})_x\|^2$

$$\left[\frac{1}{(1-t)^2}\left(\|\nabla(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\|^2 + \|\nabla^2(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\|\right)\right.$$
$$\left.+\left(\|\nabla(d\operatorname{Exp}_{x_1})_{\operatorname{Log}_{x_1}(x)}\|^2 + \|\nabla^2(d\operatorname{Exp}_{x_1})_{\operatorname{Log}_{x_1}(x)}\|\right)\right]$$
$$+ d\,\|\nabla(d\operatorname{Log}_{x_1})_x\|\left(\frac{1}{1-t}\|\nabla(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\| + \|\nabla(d\operatorname{Exp}_{x_1})_{\operatorname{Log}_{x_1}(x)}\|\right),$$

$$|\partial_t \log J_t(x\mid x_1)| \le \frac{d}{1-t} + \frac{d}{(1-t)^2}\|\operatorname{Log}_{x_1}(x)\|\,\|\nabla(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\|,$$

$\|\partial_t \text{梯度}_x \text{对数映射}\ J_t(x\mid x_1)\| \le \|(d\text{对数映射}_{x_1})_x\|\left[\frac{d}{1-t}()_2\|\nabla(d\text{指数映射}_{x_1})_{\frac{1}{1-t}\text{对数映射}_{x_1}(x)}\|\right.$

$$\left.+\frac{d}{(1-t)^3}\|\operatorname{Log}_{x_1}(x)\|\left(\|\nabla(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\|^2 + \|\nabla^2(d\operatorname{Exp}_{x_1})_{\frac{1}{1-t}\operatorname{Log}_{x_1}(x)}\|\right)\right],$$

证明至此完成。 □

E.4 界定 $\left\|\left(\nabla^2 d\,\text{指数映射}_p\right)_{rv}\right\|_{\text{op}}$

本节将给出指数映射的三阶导数上界。我们首先回顾 Lezcano-Casado (2020) 中的一些记号，这些记号可以将三阶导数表示为一个常微分方程。然后，我们应用比较定理推导上界。值得注意的是，在对称空间上，曲率张量的（协变）导数恒为零，例如参见（Lee，2018，定理 10.19）。本节中的技巧适用于非对称空间，但由于我们研究的是对称正定矩阵（n），我们假设 M 是对称空间，这将简化我们的计算。

回忆一下，∇ 表示 M 上的 TM 的（Levi–Civita）联络。令 ∇^{flat} 为向量空间 T_pM 上的平坦联络。在 p 的正规邻域上定义指数映射 $\exp_p\colon U \subset T_pM \to M$。

我们首先引入拉回联络。除了沿着 M上的向量场求导，我们也可以沿着 T_pM上的向量场 X 求导。将相应的联络定义为

$$\big(\nabla^{\text{pullback}}_X Y\big)(u) = \nabla_{(d\operatorname{Exp}_p)_u(X(u))} Y,$$

换言之，给定 $u \in T_pM$，且有 $q := $指数映射 $_p(u)$。那么，通过拉回联络沿切空间中的 X 求导，并在 $u \in T_pM$ 处取值，等价于沿 d指数映射 $_p(X)$ 求导（可将其视为定义在 M 上的向量场），并在 q 处取值。

回忆一下，拉回丛 $\operatorname{Exp}^*_p(TM)$ 是基空间 T_pM 上的向量丛。具体而言，

$$\operatorname{Exp}^*_p(TM) := \{(u, v) : u \in T_pM, v \in TM \text{ with base point } \operatorname{Exp}_p(u)\}.$$

例如，给定张量积空间 $(\alpha, (u, v)) \in T^*_pM \otimes \operatorname{Exp}^*_p(TM)$ 的一个元素，我们有 $(\alpha, (u, v))(w) = \alpha(w)v$, 其中 $\alpha \in T^*_pM$ 是一个余向量，因此 $\alpha(w) \in \mathbb{R}$。

令 $\{e_i\}$ 为 T_pM 的一组正交归一基，$\{e^i\}$ 为余切空间 T^*_pM 中对应的基。根据定义，

$$\big(\sum_i e^i \otimes (u, (d\operatorname{Exp}_p)_u[e_i])\big)(w) := \sum_i e^i(w)(d\operatorname{Exp}_p)_u[e_i],$$

并利用线性性

$$\sum_i e^i(w)(d\operatorname{Exp}_p)_u[e_i] = (d\operatorname{Exp}_p)_u[\sum_i e^i(w)e_i] = (d\operatorname{Exp}_p)_u(w).$$

由此我们看到，$d\operatorname{Exp}_p$ 可以视为张量积空间中的一个元素：

$$(d\operatorname{Exp}_p)_u \sum_i e^i \otimes (u, (d\operatorname{Exp}_p)_u[e_i]) \in T^*_pM \otimes \operatorname{Exp}^*_p(TM).$$

我们赋予 T^*_pM由 ∇^{flat} 诱导的联络，并赋予 $\operatorname{Exp}^*_p(TM)$ ∇^{pullback}。这些联络在张量积丛上诱导出一个联络，记为 ∇^{induced}，其特征由莱布尼茨法则给出。现在考虑 W_1, W_2为切空间 T_pM中的向量场，而 $d\operatorname{Exp}_p(W_2)$可视为一个截面，即 $d\operatorname{Exp}_p(W_2) : u \mapsto (d\operatorname{Exp}_p)_u(W_2(u))$。我们对其求导

$$\begin{aligned}
&\nabla^{\text{pullback}}_{W_1}(d\operatorname{Exp}_p(W_2)) = \nabla^{\text{pullback}}_{W_1}(\sum_i e^i(W_2)(d\operatorname{Exp}_p)[e_i]) \\
=&\sum_i W_1 e^i(W_2)(d\operatorname{Exp}_p)[e_i] + \sum_i e^i(W_2)\nabla^{\text{pullback}}_{W_1}(d\operatorname{Exp}_p)[e_i].
\end{aligned}$$

这里我们强调，$(d\operatorname{Exp}_p)[e_i]$ 是 u的函数，与之前类似：$(d\operatorname{Exp}_p)[e_i] : u \mapsto (d\operatorname{Exp}_p)_u[e_i]$。

由$(d\operatorname{Exp}_p)$的线性性以及平坦联络的定义，我们有

$$\sum_i W_1 e^i(W_2)(d\operatorname{Exp}_p)[e_i] = (d\operatorname{Exp}_p)[\sum_i W_1 e^i(W_2)e_i] = (d\operatorname{Exp}_p)[\nabla^{\text{flat}}_{W_1} W_2].$$

另一方面，由于 e^i 是常数，我们知道 $\nabla^{\text{flat}}_{W_1} e^i = 0$，因此

$$e^i \otimes \nabla^{\text{pullback}}_{W_1}(d\operatorname{Exp}_p)[e_i] = e^i \otimes \nabla^{\text{pullback}}_{W_1}(d\operatorname{Exp}_p)[e_i] + \nabla^{\text{flat}}_{W_1} e^i \otimes (d\operatorname{Exp}_p)[e_i]$$

$$=\nabla^{\text{induced}}_{\overline{W}_1}(e^i\otimes(d\operatorname{Exp}_p)[e_i]),$$

因此

$$\begin{aligned}&\sum_i e^i(\overline{W}_2)\otimes\nabla^{\text{pullback}}_{\overline{W}_1}((d\operatorname{Exp}_p)[e_i])=\Big(\sum_i e^i\otimes\nabla^{\text{pullback}}_{\overline{W}_1}((d\operatorname{Exp}_p)[e_i])\Big)(\overline{W}_2)\\ =&\Big(\sum_i\nabla^{\text{induced}}_{\overline{W}_1}(e^i\otimes(d\operatorname{Exp}_p)[e_i])\Big)(\overline{W}_2)=\sum_i\Big(\nabla^{\text{induced}}_{\overline{W}_1}(e^i\otimes(d\operatorname{Exp}_p)[e_i])\Big)(\overline{W}_2)\\ =&(\nabla^{\text{induced}}_{\overline{W}_1}(\sum_i e^i\otimes(d\operatorname{Exp}_p)[e_i]))(\overline{W}_2)=(\nabla^{\text{induced}}_{\overline{W}_1}d\operatorname{Exp}_p)(\overline{W}_2).\end{aligned}$$

代入上一个表达式，我们得到

$$\begin{aligned}&\nabla^{\text{pullback}}_{\overline{W}_1}(d\operatorname{Exp}_p(\overline{W}_2))\\ =&\sum_i\overline{W}_1e^i(\overline{W}_2)(d\operatorname{Exp}_p)[e_i]+\sum_i e^i(\overline{W}_2)\nabla^{\text{pullback}}_{\overline{W}_1}(d\operatorname{Exp}_p)[e_i]\\ =&(d\operatorname{Exp}_p)[\nabla^{\text{flat}}_{\overline{W}_1}\overline{W}_2]+\nabla^{\text{induced}}_{\overline{W}_1}(d\operatorname{Exp}_p)(\overline{W}_2)\end{aligned}$$

因此，对于向量场 $\overline{W_1},\overline{W_2}$ 在 $U\subset T_pM$ 上，有

$$\nabla^{\text{pullback}}_{\overline{W}_1}\big(d\operatorname{Exp}_p(\overline{W}_2)\big)=\big(\nabla^{\text{induced}}_{\overline{W}_1}d\operatorname{Exp}_p\big)(\overline{W}_2)+d\operatorname{Exp}_p\big(\nabla^{\text{flat}}_{\overline{W}_1}\overline{W}_2\big).\tag{27}$$

定义 $c(t,s_1,s_2,s_3)=$指数映射$_p(t(v+s_1w_1+s_2w_2+s_3w_3))$，以及 $\gamma(t)=$指数映射$_p(tv)$。我们简要回顾 Lezcano-Casado（2020）为控制指数映射的二阶导数所做的工作。定义$J_1(t)=(d$指数映射$_p)_{tv}(tw_1)$为沿着 γ的雅可比场，其初始条件为 w_1。我们可以定义$\widetilde{}$在 w_2方向上对 J_1的延拓为 $J_1(t,s):=(d$指数映射$_p)_{t(v+sw_2)}(tw_1)$，注意$\sim J_1(t,0)=J_1(t)$。对于固定的 t，切空间向量场 $u\mapsto tw_1$ 在 T_pM上是常数，因此

$$\nabla^{\text{flat}}_{tw_2}(tw_1)=0.\tag{28}$$

将（27）应用于 $\overline{W_1}\equiv tw_2$和 $\overline{W_2}\equiv tw_1$，并在 $t(v+sw_2)$ 处求值，可得

$$\big(\nabla^{\text{pullback}}_{tw_2}\widetilde{J}_1(t,s)\big)(t(v+sw_2))=\big(\nabla^{\text{induced}}_{tw_2}d\operatorname{Exp}_p\big)_{t(v+sw_2)}(tw_1)+(d\operatorname{Exp}_p)_{t(v+sw_2)}\big(\nabla^{\text{flat}}_{tw_2}(tw_1)\big).$$

根据（28），最后一项为零。利用这一事实 $(\nabla^{\text{pullback}}_XY)(u)=\nabla_{(d\text{指数映射}_p)_u(X(u))}Y$其中 $X=tw_2,\sim Y=J_1,u=tv$，可得（限制在 $s=0$上）

$$\big(\nabla^{\text{pullback}}_{tw_2}\widetilde{J}_1\big)(tv)=\nabla_{(d\operatorname{Exp}_p)_{tv}(tw_2)}\widetilde{J}_1.$$

结合先前定义的记号 $J_2=(d$指数映射$_p)_{tv}(tw_2)$，我们有 e

$$\nabla_{J_2}J_1=\nabla_{J_2}\widetilde{J}_1\big|_{s=0}=\nabla^{\text{pullback}}_{tw_2}\widetilde{J}_1\big|_{s=0}=\big(\nabla^{\text{induced}}d\operatorname{Exp}_p\big)_{tv}(tw_1,tw_2).$$

因此，我们将沿着 γ 的二阶变分场定义为

$$K_{12}(t):=\big(\nabla^{\text{induced}}d\operatorname{Exp}_p\big)_{tv}(tw_1,tw_2)=\nabla_{J_2}J_1.$$

那么（Lezcano-Casado，2020，命题 4.1）表明 K_{12} 满足

$$\ddot{K}_{12}+R(K_{12},\dot{\gamma})\dot{\gamma}+Y_{12}=0,\qquad K_{12}(0)=0,\ \dot{K}_{12}(0)=0,\tag{29}$$

其中 Y_{12} 在该文中有明确给出的表达式。采用相同的技巧，我们可以定义

$$K_{13}(t) := \big(\nabla^{\text{induced}} d\operatorname{Exp}_p\big)_{tv}(tw_1, tw_3) = \nabla_{J_3} J_1, K_{23}(t) := \big(\nabla^{\text{induced}} d\operatorname{Exp}_p\big)_{tv}(tw_2, tw_3) = \nabla_{J_3} J_2,$$

满足相应的常微分方程，如（29）。

现在我们研究三阶导数 $L_{123} = \big((\nabla^{\text{induced}})^2 d\text{指数映射}_p\big)(tw_1, tw_2, tw_3)$。定义 $J_i(t)$ 为沿着 γ 的雅可比场，其初始条件为 w_i，其中 $i = 2, 3$；定义方式与之前对 J_1 的处理相同。将 $K_{12}(t) = \big(\nabla^{\text{induced}} d\text{指数映射}_p\big)_{tv}(tw_1, tw_2) = \nabla_{J_2} J_1$ 从 γ 延拓到其外部：

$$\widetilde{K}_{12}(t, s_3) := \big(\nabla^{\text{induced}} d\operatorname{Exp}_p\big)_{t(v+s_3 w_3)}(tw_1, tw_2),$$

使得 $\widetilde{\ }$ 在 $K_{12}(t, 0) = K_{12}(t)$。随后，我们定义沿着 γ 的三阶变分场：

$$L_{123}(t) := \nabla_{J_3} K_{12}(t) := \nabla_{J_3} \widetilde{K}_{12}(t, s_3)\Big|_{s_3=0}. \tag{30}$$

将 $\nabla^{\text{pullback}}_{W_3 —}$ 应用于一阶莱布尼茨法则（27）。对于 W_1, W_2, W_3 上的向量场 T_pM，我们在任意 $u \in T_pM$ 处得到，

$$\begin{aligned}\big(\nabla^{\text{pullback}}_{\overline{W}_3} \nabla^{\text{pullback}}_{\overline{W}_1} (d\operatorname{Exp}_p(\overline{W}_2))\big)(u) = {} & \big((\nabla^{\text{induced}}_{\overline{W}_3} \nabla^{\text{induced}}_{\overline{W}_1} d\operatorname{Exp}_p)_u\big)(\overline{W}_2(u)) \\ & + \big(\nabla^{\text{induced}}_{\overline{W}_1} d\operatorname{Exp}_p\big)_u\big((\nabla^{\text{flat}}_{\overline{W}_3} \overline{W}_2)(u)\big) \\ & + \big(\nabla^{\text{induced}}_{\overline{W}_3} d\operatorname{Exp}_p\big)_u\big((\nabla^{\text{flat}}_{\overline{W}_1} \overline{W}_2)(u)\big) \\ & + (d\operatorname{Exp}_p)_u\big((\nabla^{\text{flat}}_{\overline{W}_3} \nabla^{\text{flat}}_{\overline{W}_1} \overline{W}_2)(u)\big).\end{aligned} \tag{31}$$

现在将（31）特化到常向量场 $W_1 \equiv tw_2$, $W_2 \equiv \overline{tw_1}$, $W_3 \equiv \overline{tw_3}$。由于 $\overline{\nabla}^{\text{flat}}_{tw_i}(tw_j) = 0$ 对所有 i, j，所以（31）中的最后三项为零。在 $u = tv$ 处求值得到

$$L_{123}(t) = \big(\nabla^{\text{pullback}}_{tw_3} \widetilde{K}_{12}\big)(tv) = \big((\nabla^{\text{induced}}_{tw_3} \nabla^{\text{induced}}_{tw_2} d\operatorname{Exp}_p)_{tv}\big)(tw_1),$$

其中注意到 $(d\text{指数映射}_p)_{tv}(tw_3) = J_3(t)$。

根据定义 二阶协变导数张量 $(\nabla^{\text{induced}})^2 d\text{指数映射}_p$ 的定义，这恰好是

$$L_{123}(t) = \big((\nabla^{\text{induced}})^2 d\operatorname{Exp}_p\big)_{tv}(tw_1, tw_2, tw_3). \tag{32}$$

在本节中，我们考虑沿 J_i 的协变导数，例如 $\nabla_{J_3} J_2$，作为 $\widetilde{\ } \nabla_{J_3} J_2$ 的缩写。这里，$\widetilde{\ }$ 表示沿适当方向的相应延拓，如上所构造。

引理 E.9. 对于上面定义的 L_{123}，我们有

$$\ddot{L}_{123} + R(L_{123}, \dot{\gamma})\dot{\gamma} + Y_{123} = 0, \qquad L_{123}(0) = 0,\ \dot{L}_{123}(0) = 0.$$

当 M 是对称空间时，

$$\begin{aligned}Y_{123} = {} & R(\dot{J}_3, \dot{\gamma}) K_{12} + 2R(J_3, \dot{\gamma}) \dot{K}_{12} + R(K_{12}, \dot{J}_3)\dot{\gamma} + R(K_{12}, \dot{\gamma})\dot{J}_3 \\ & + 2R(\nabla_{J_3} J_1, \dot{\gamma})\dot{J}_2 + 2R(J_1, \dot{J}_3)\dot{J}_2 + 2R(J_1, \dot{\gamma})\nabla_{J_3}\dot{J}_2 \\ & + 2R(\nabla_{J_3} J_2, \dot{\gamma})\dot{J}_1 + 2R(J_2, \dot{J}_3)\dot{J}_1 + 2R(J_2, \dot{\gamma})\nabla_{J_3}\dot{J}_1.\end{aligned}$$

证明。 现在回顾一下，下列常微分方程由 K （Lezcano-Casado，2020，命题 4.1。）

$$\ddot{K}_{12} + R(K_{12}, \dot{\gamma})\dot{\gamma} + Y_{12} = 0.$$

对常微分方程求导（注意，我们取协变导数所针对的向量场是 $\tilde{}$ 视为相应的延拓，即 K_{12} 对于 K_{12}，梯度 r 对于 $\dot{\gamma}$，以及 $\nabla_{\operatorname{grad} r}\nabla_{\operatorname{grad} r}K_{12}$ 对于 $\ddot{}K_{12}$），我们得到

$$\nabla_{J_3}\ddot{K}_{12} + \nabla_{J_3}R(K_{12}, \dot{\gamma})\dot{\gamma} + \nabla_{J_3}Y_{12} = 0,$$

$\ddot{}$ $\dot{}$ $\ddot{}$我们首先计算 $\nabla_{J_3}K_{12}$。回顾一下，$X := \nabla_{\dot{\gamma}}X$ 以及 $X := \nabla_{\dot{\gamma}}\nabla_{\dot{\gamma}}X$。令 $L_{123} := \nabla_{J_3}K_{12}$。我们逐步计算 $\nabla_{J_3}K_{12}$。

根据曲率张量的定义，回顾

$$\nabla_{J_3}\nabla_{\dot{\gamma}}X = \nabla_{\dot{\gamma}}\nabla_{J_3}X + \nabla_{[J_3,\dot{\gamma}]}X + R(J_3, \dot{\gamma})X,$$

以及以下事实 $[J_3, \dot{\gamma}] = 0$ （理由与Lezcano-Casado （2020，命题 4.1）中的相同），我们有

$$\begin{aligned}\nabla_{J_3}\ddot{K}_{12} &= \nabla_{J_3}\nabla_{\dot{\gamma}}\dot{K}_{12} = \nabla_{\dot{\gamma}}\nabla_{J_3}\dot{K}_{12} + R(J_3, \dot{\gamma})\dot{K}_{12}\\ &= \nabla_{\dot{\gamma}}\nabla_{J_3}\nabla_{\dot{\gamma}}K_{12} + R(J_3, \dot{\gamma})\dot{K}_{12}\\ &= \nabla_{\dot{\gamma}}(\nabla_{\dot{\gamma}}\nabla_{J_3}K_{12} + R(J_3, \dot{\gamma})K_{12}) + R(J_3, \dot{\gamma})\dot{K}_{12} = \nabla_{\dot{\gamma}}\nabla_{\dot{\gamma}}L_{123} + \nabla_{\dot{\gamma}}(R(J_3, \dot{\gamma})K_{12}) + R(J_3, \dot{\gamma})\dot{K}_{12}.\end{aligned}$$

注意到（根据莱布尼茨法则）

$$\begin{aligned}\nabla_{\dot{\gamma}}(R(J_3, \dot{\gamma})K_{12}) &= (\nabla_{\dot{\gamma}}R)(J_3, \dot{\gamma})K_{12} + R(\nabla_{\dot{\gamma}}J_3, \dot{\gamma})K_{12} + R(J_3, \nabla_{\dot{\gamma}}\dot{\gamma})K_{12} + R(J_3, \dot{\gamma})\nabla_{\dot{\gamma}}K_{12}\\ &= (\nabla_{\dot{\gamma}}R)(J_3, \dot{\gamma})K_{12} + R(\dot{J}_3, \dot{\gamma})K_{12} + R(J_3, \dot{\gamma})\dot{K}_{12}.\end{aligned}$$

因此我们有

$$\nabla_{J_3}\ddot{K}_{12} = \ddot{L}_{123} + (\nabla_{\dot{\gamma}}R)(J_3, \dot{\gamma})K_{12} + R(\dot{J}_3, \dot{\gamma})K_{12} + 2R(J_3, \dot{\gamma})\dot{K}_{12}.$$

现在我们计算 $\nabla_{J_3}\big(R(K_{12}, \dot{\gamma})\dot{\gamma}\big)$。注意到 $\nabla_{J_3}\dot{\gamma} = \nabla_{\dot{\gamma}}J_3 = \dot{J}_3$，根据莱布尼茨法则，

$$\begin{aligned}\nabla_{J_3}\big(R(K_{12}, \dot{\gamma})\dot{\gamma}\big) &= (\nabla_{J_3}R)(K_{12}, \dot{\gamma})\dot{\gamma} + R(\nabla_{J_3}K_{12}, \dot{\gamma})\dot{\gamma} + R(K_{12}, \nabla_{J_3}\dot{\gamma})\dot{\gamma} + R(K_{12}, \dot{\gamma})\nabla_{J_3}\dot{\gamma}\\ &= (\nabla_{J_3}R)(K_{12}, \dot{\gamma})\dot{\gamma} + R(L_{123}, \dot{\gamma})\dot{\gamma} + R(K_{12}, \dot{J}_3)\dot{\gamma} + R(K_{12}, \dot{\gamma})\dot{J}_3.\end{aligned}$$

回忆

$$Y := 2R(J_1, \dot{\gamma})\dot{J}_2 + 2R(J_2, \dot{\gamma})\dot{J}_1 + (\nabla_{\dot{\gamma}}R)(J_2, \dot{\gamma})J_1 + (\nabla_{J_2}R)(J_1, \dot{\gamma})\dot{\gamma}.$$

$\dot{}$另外回忆 $[J_3, \dot{\gamma}] = 0$ 和 $\nabla_{J_3}\dot{\gamma} = \nabla_{\dot{\gamma}}J_3 = J_3$。然后，根据张量场的莱布尼茨法则，我们有

$$\begin{aligned}\nabla_{J_3}Y = {} & 2\,\nabla_{J_3}\big(R(J_1, \dot{\gamma})\dot{J}_2\big) + 2\,\nabla_{J_3}\big(R(J_2, \dot{\gamma})\dot{J}_1\big)\\ & + \nabla_{J_3}\Big((\nabla_{\dot{\gamma}}R)(J_2, \dot{\gamma})J_1\Big) + \nabla_{J_3}\Big((\nabla_{J_2}R)(J_1, \dot{\gamma})\dot{\gamma}\Big),\end{aligned}$$

其中各项展开如下。对于第一项，

$$\begin{aligned}\nabla_{J_3}\big(R(J_1, \dot{\gamma})\dot{J}_2\big) &= (\nabla_{J_3}R)(J_1, \dot{\gamma})\dot{J}_2 + R(\nabla_{J_3}J_1, \dot{\gamma})\dot{J}_2 + R(J_1, \nabla_{J_3}\dot{\gamma})\dot{J}_2 + R(J_1, \dot{\gamma})\nabla_{J_3}\dot{J}_2\\ &= (\nabla_{J_3}R)(J_1, \dot{\gamma})\dot{J}_2 + R(\nabla_{J_3}J_1, \dot{\gamma})\dot{J}_2 + R(J_1, \dot{J}_3)\dot{J}_2 + R(J_1, \dot{\gamma})\nabla_{J_3}\dot{J}_2.\end{aligned}$$

对于第二项,

$$\begin{aligned}\nabla_{J_3}\big(R(J_2,\dot\gamma)\dot J_1\big) &= (\nabla_{J_3}R)(J_2,\dot\gamma)\dot J_1 + R(\nabla_{J_3}J_2,\dot\gamma)\dot J_1 + R(J_2,\nabla_{J_3}\dot\gamma)\dot J_1 + R(J_2,\dot\gamma)\nabla_{J_3}\dot J_1\\ &= (\nabla_{J_3}R)(J_2,\dot\gamma)\dot J_1 + R(\nabla_{J_3}J_2,\dot\gamma)\dot J_1 + R(J_2,\dot J_3)\dot J_1 + R(J_2,\dot\gamma)\nabla_{J_3}\dot J_1.\end{aligned}$$

对于第三项,

$$\begin{aligned}\nabla_{J_3}\Big((\nabla_{\dot\gamma}R)(J_2,\dot\gamma)J_1\Big) &= (\nabla_{J_3}\nabla_{\dot\gamma}R)(J_2,\dot\gamma)J_1 + (\nabla_{\dot\gamma}R)(\nabla_{J_3}J_2,\dot\gamma)J_1\\ &\quad + (\nabla_{\dot\gamma}R)(J_2,\nabla_{J_3}\dot\gamma)J_1 + (\nabla_{\dot\gamma}R)(J_2,\dot\gamma)\nabla_{J_3}J_1\\ &= (\nabla_{J_3}\nabla_{\dot\gamma}R)(J_2,\dot\gamma)J_1 + (\nabla_{\dot\gamma}R)(\nabla_{J_3}J_2,\dot\gamma)J_1\\ &\quad + (\nabla_{\dot\gamma}R)(J_2,\dot J_3)J_1 + (\nabla_{\dot\gamma}R)(J_2,\dot\gamma)\nabla_{J_3}J_1.\end{aligned}$$

对于最后一项,

$$\begin{aligned}\nabla_{J_3}\Big((\nabla_{J_2}R)(J_1,\dot\gamma)\dot\gamma\Big) &= (\nabla_{J_3}\nabla_{J_2}R)(J_1,\dot\gamma)\dot\gamma + (\nabla_{J_2}R)(\nabla_{J_3}J_1,\dot\gamma)\dot\gamma\\ &\quad + (\nabla_{J_2}R)(J_1,\nabla_{J_3}\dot\gamma)\dot\gamma + (\nabla_{J_2}R)(J_1,\dot\gamma)\nabla_{J_3}\dot\gamma\\ &= (\nabla_{J_3}\nabla_{J_2}R)(J_1,\dot\gamma)\dot\gamma + (\nabla_{J_2}R)(\nabla_{J_3}J_1,\dot\gamma)\dot\gamma\\ &\quad + (\nabla_{J_2}R)(J_1,\dot J_3)\dot\gamma + (\nabla_{J_2}R)(J_1,\dot\gamma)\dot J_3.\end{aligned}$$

代入各表达式，得到

$$\ddot L_{123} + R(L_{123},\dot\gamma)\dot\gamma + Y_{123} = 0,$$

其中

$$\begin{aligned}Y_{123} &= (\nabla_{\dot\gamma}R)(J_3,\dot\gamma)K_{12} + (\nabla_{J_3}R)(K_{12},\dot\gamma)\dot\gamma + R(\dot J_3,\dot\gamma)K_{12} + 2R(J_3,\dot\gamma)\dot K_{12}\\ &\quad + R(K_{12},\dot J_3)\dot\gamma + R(K_{12},\dot\gamma)\dot J_3\\ &\quad + 2(\nabla_{J_3}R)(J_1,\dot\gamma)\dot J_2 + 2R(\nabla_{J_3}J_1,\dot\gamma)\dot J_2 + 2R(J_1,\dot J_3)\dot J_2 + 2R(J_1,\dot\gamma)\nabla_{J_3}\dot J_2\\ &\quad + 2(\nabla_{J_3}R)(J_2,\dot\gamma)\dot J_1 + 2R(\nabla_{J_3}J_2,\dot\gamma)\dot J_1 + 2R(J_2,\dot J_3)\dot J_1 + 2R(J_2,\dot\gamma)\nabla_{J_3}\dot J_1\\ &\quad + (\nabla_{J_3}\nabla_{\dot\gamma}R)(J_2,\dot\gamma)J_1 + (\nabla_{\dot\gamma}R)(\nabla_{J_3}J_2,\dot\gamma)J_1 + (\nabla_{\dot\gamma}R)(J_2,\dot J_3)J_1 + (\nabla_{\dot\gamma}R)(J_2,\dot\gamma)\nabla_{J_3}J_1\\ &\quad + (\nabla_{J_3}\nabla_{J_2}R)(J_1,\dot\gamma)\dot\gamma + (\nabla_{J_2}R)(\nabla_{J_3}J_1,\dot\gamma)\dot\gamma + (\nabla_{J_2}R)(J_1,\dot J_3)\dot\gamma + (\nabla_{J_2}R)(J_1,\dot\gamma)\dot J_3.\end{aligned}$$

在对称空间上,

$$\begin{aligned}Y_{123} &= R(\dot J_3,\dot\gamma)K_{12} + 2R(J_3,\dot\gamma)\dot K_{12} + R(K_{12},\dot J_3)\dot\gamma + R(K_{12},\dot\gamma)\dot J_3\\ &\quad + 2R(\nabla_{J_3}J_1,\dot\gamma)\dot J_2 + 2R(J_1,\dot J_3)\dot J_2 + 2R(J_1,\dot\gamma)\nabla_{J_3}\dot J_2\\ &\quad + 2R(\nabla_{J_3}J_2,\dot\gamma)\dot J_1 + 2R(J_2,\dot J_3)\dot J_1 + 2R(J_2,\dot\gamma)\nabla_{J_3}\dot J_1.\end{aligned}$$

□

现在我们可以应 比较理论（Lezcano-Casado， 2020， 命题4.9）来控制 $\|L\|$.

引理 E.10。 设 (M,g) 是一个黎曼 对称空间，即 $\nabla R\equiv 0$。固定一条单位速度测地线 $\gamma\colon\ [0,r]\to M$ 满足 $\|\dot\gamma\|\equiv 1$。在假设 *1*下，我们有

$$\|L(t)\| \le \frac{7L_R^2}{3}s_{K_{\min}}(s)^5 + 16L_R^3t^2s_{K_{\min}}(t)^5 + 20L_R^2ts_{K_{\min}}(t)^4.$$

因此，

$$\left\|\left((\nabla)^2 d\operatorname{Exp}_p\right)_{rv}\right\|_{\mathrm{op}} = \sup_{\|w_1\|,\|w_2\|,\|w_3\|\le 1}\left\|\left((\nabla^{induced})^2 d\operatorname{Exp}_p\right)_{rv}(w_1,w_2,w_3)\right\|$$
$$= \sup_{\|w_1\|,\|w_2\|,\|w_3\|\le 1}\frac{1}{r^3}\|L_{123}(r)\| \le \frac{7L_R^2}{3r^3}s_{K_{\min}}(r)^5 + 16L_R^3\frac{1}{r}s_{K_{\min}}(r)^5 + 20L_R^2\frac{1}{r^2}s_{K_{\min}}(r)^4.$$

证明。回顾

$$\begin{aligned}Y_{123} = {} & R(\dot{J}_3,\dot{\gamma})K_{12} + 2R(J_3,\dot{\gamma})\dot{K}_{12} + R(K_{12},\dot{J}_3)\dot{\gamma} + R(K_{12},\dot{\gamma})\dot{J}_3\\ & + 2R(\nabla_{J_3}J_1,\dot{\gamma})\dot{J}_2 + 2R(J_1,\dot{J}_3)\dot{J}_2 + 2R(J_1,\dot{\gamma})\nabla_{J_3}\dot{J}_2\\ & + 2R(\nabla_{J_3}J_2,\dot{\gamma})\dot{J}_1 + 2R(J_2,\dot{J}_3)\dot{J}_1 + 2R(J_2,\dot{\gamma})\nabla_{J_3}\dot{J}_1.\end{aligned}$$

根据 K_{13}, K_{23}的定义，我们有 $\nabla_{J_3}J_1 = K_{13}$ 和 $\nabla_{J_3}J_2 = K_{23}$。接下来，我们用 K_{i3} 和曲率来表示 $\nabla_{J_3}\dot{J}_i$。根据 R 和 $[J_3,\dot{\gamma}] = 0$的定义，

$$\begin{aligned}\nabla_{J_3}\dot{J}_i = \nabla_{J_3}\nabla_{\dot{\gamma}}J_i &= \nabla_{\dot{\gamma}}\nabla_{J_3}J_i + R(J_3,\dot{\gamma})J_i\\ &= \dot{K}_{i3} + R(J_3,\dot{\gamma})J_i.\end{aligned}$$

因此

$$\begin{aligned}2R(J_1,\dot{\gamma})\nabla_{J_3}\dot{J}_2 &= 2R(J_1,\dot{\gamma})\dot{K}_{23} + 2R(J_1,\dot{\gamma})\big(R(J_3,\dot{\gamma})J_2\big),\\ 2R(J_2,\dot{\gamma})\nabla_{J_3}\dot{J}_1 &= 2R(J_2,\dot{\gamma})\dot{K}_{13} + 2R(J_2,\dot{\gamma})\big(R(J_3,\dot{\gamma})J_1\big).\end{aligned}$$

将这些结果代入并利用 $R(x,y)z \le L_R\|x\|\|y\|\|z\|$，我们有

$$\begin{aligned}\|Y_{123}\| \le {} & 3L_R\|K_{12}\|\|\dot{J}_3\| + 2L_R\|J_3\|\|\dot{K}_{12}\| + 2L_R\|K_{13}\|\|\dot{J}_2\| + 2L_R\|K_{23}\|\|\dot{J}_1\|\\ & + 2L_R\|J_1\|\|\dot{K}_{23}\| + 2L_R\|J_2\|\|\dot{K}_{13}\| + 2L_R\|J_1\|\|\dot{J}_2\|\|\dot{J}_3\| + 2L_R\|J_2\|\|\dot{J}_1\|\|\dot{J}_3\|\\ & + 4L_R^2\|J_1\|\|J_2\|\|J_3\|\\ \le {} & 7L_R\|K\|\|\dot{J}\| + 6L_R\|J\|\|\dot{K}\| + 4L_R\|J\|\|\dot{J}\|^2 + 4L_R^2\|J\|^3.\end{aligned}$$

Usin引理E.11,(Lezcano-Casado, 2020, 定理 3.12, 定理 4.11），我们有

$$\begin{aligned}\|K(t)\| &\le \frac{16}{3}s_{K_{\min}}(t/2)^2 L_R s_{K_{\min}}(t),\\ \|\dot{K}(t)\| &\le L_R^2\frac{16}{3}t s_{K_{\min}}(t)^3 + 2L_R s_{K_{\min}}(t)^2,\\ \|J(t)\| &\le s_{K_{\min}}(t),\\ \|\dot{J}(t)\| &\le s'_{K_{\min}}(t),\end{aligned}$$

因此我们可以给出上界

$$\begin{aligned}\|Y_{123}\| \le {} & 7L_R\frac{16}{3}s_{K_{\min}}(t/2)^2 L_R s_{K_{\min}}(t)s'_{K_{\min}}(t) + 6L_R s_{K_{\min}}(t)(L_R^2\frac{16}{3}t s_{K_{\min}}(t)^3 + 2L_R s_{K_{\min}}(t)^2)\\ & + 4L_R s_{K_{\min}}(t)(s'_{K_{\min}}(t))^2 + 4L_R^2(s_{K_{\min}}(t))^3\\ \le {} & \frac{112L_R^2}{3}s_{K_{\min}}(t/2)^2 s_{K_{\min}}(t)s'_{K_{\min}}(t) + 32L_R^3 t s_{K_{\min}}(t)^4 + 20L_R^2 s_{K_{\min}}(t)^3\\ \le {} & \frac{28L_R^2}{3}s_{K_{\min}}(t)^3 s'_{K_{\min}}(t) + 32L_R^3 t s_{K_{\min}}(t)^4 + 20L_R^2 s_{K_{\min}}(t)^3.\end{aligned}$$

我们应用（Lezcano-Casado，2020，命题4.9）。考虑常微分方程

$$\rho''(t) + K_{\min}\rho(t) = \eta(t).$$

定义 $\int_0^t y\ t-s\ \eta\ s\ ds$ $= s_{K_{\min}}$ y $\rho''\ t\ + K_{\min}\rho\ t\ = \eta\ t$ ρ $= Y_{123}\ (t)\ (t)$。注意，解满足 $y''(t)\ + K_{\min}y(t)\ = 0$。于是，函数 $(t)\ (\ )\ (\ )$ 满足 $(\ )\ (\ )\ (\ )$。我们将此应用于 上界的各个部分：

$$\int_0^t s_{K_{\min}}(t-s)\frac{28L_R^2}{3}s_{K_{\min}}(s)^3 s'_{K_{\min}}(s)ds \le \frac{7L_R^2}{3}s_{K_{\min}}(t)\int_0^t 4s_{K_{\min}}(s)^3 s'_{K_{\min}}(s)ds = \frac{7L_R^2}{3}s_{K_{\min}}(s)^5,$$

$$\begin{aligned}\int_0^t s_{K_{\min}}(t-s)32L_R^3 s_{K_{\min}}(s)^4 sds &\le 32L_R^3 s_{K_{\min}}(t)\int_0^t s_{K_{\min}}(s)^4 sds\\ &\le 32L_R^3 s_{K_{\min}}(t)^5\int_0^t sds = 16L_R^3 t^2 s_{K_{\min}}(t)^5,\end{aligned}$$

$$\int_0^t s_{K_{\min}}(t-s)20L_R^2 s_{K_{\min}}(s)^3 ds \le 20L_R^2 t s_{K_{\min}}(t)^4.$$

因此，我们可以使用以下 $\tilde{\rho}$ 作为 ρ的上界：

$$\rho(t) \le \tilde{\rho}(t) := \frac{7L_R^2}{3}s_{K_{\min}}(s)^5 + 16L_R^3 t^2 s_{K_{\min}}(t)^5 + 20L_R^2 t s_{K_{\min}}(t)^4.$$

换句话说，

$$\|L(t)\| \le \frac{7L_R^2}{3}s_{K_{\min}}(s)^5 + 16L_R^3 t^2 s_{K_{\min}}(t)^5 + 20L_R^2 t s_{K_{\min}}(t)^4.$$

□

引理 E.11。 当 M 是对称空间时，我们有

$$\|\dot{K}(t)\| \le L_R^2\frac{16}{3}t s_{K_{\min}}(t)^3 + 2L_R s_{K_{\min}}(t)^2.$$

证明。 由于 K 满足 $\ddot{K} + R(K,\dot{\gamma})\dot{\gamma} + Y = 0$，因此我们有

$$\dddot{K} = -R(K,\dot{\gamma})\ddot{\gamma} - \dot{Y}.$$

对 $[0,t]$ 积分，得到

$$\ddot{K}(t) - \ddot{K}(0) = \int_0^t \dddot{K}(s)ds = \int_0^t -R(K,\dot{\gamma})\ddot{\gamma}(s) - \dot{Y}(s)ds.$$

利用 $\|\dot{\gamma}\| = 1$ 和 $K(0) = 0$，得到

$$\|\dot{K}(t)\| = \|\int_0^t \dddot{K}(s)ds\| \le \int_0^t \|R(K,\dot{\gamma})\ddot{\gamma}(s)\| + \|\dot{Y}(s)\|ds \le \int_0^t L_R\|K(s)\| + \|\dot{Y}(s)\|ds.$$

只需得到 $\|\dot{Y}\|$ 和 $\|K\|$ 的界。回顾

$$Y := 2R(J_1,\dot{\gamma})\dot{J}_2 + 2R(J_2,\dot{\gamma})\dot{J}_1 + (\nabla_{\dot{\gamma}}R)(J_2,\dot{\gamma})J_1 + (\nabla_{J_2}R)(J_1,\dot{\gamma})\dot{\gamma},$$

并且对于对称空间，我们有回顾

$$Y = 2R(J_1, \dot{\gamma})\dot{J}_2 + 2R(J_2, \dot{\gamma})\dot{J}_1.$$

因此

$$\|Y(s)\| \leq 2L_R(\|J_1\|\|\dot{J}_2\| + \|\dot{J}_1\|\|J_2\|) \leq 4L_R s'_{K_{\min}}(s)s_{K_{\min}}(s) = 2L_R s_{K_{\min}}(2s),$$

其中，根据 Lezcano-Casado (2020，定理 3.12)，对于单位切向量 v，

$$\|d(\operatorname{Exp}_p)_{rv}(rw)\| \leq \max\{1, \frac{s_{K_{\min}}(r)}{r}\}r\|w\|,$$

从而 $\|J(s)\| \leq s_{K_{\min}}(s)$。此外，Lezcano-Casado (2020，定理 3.11) 给出

$$\|\operatorname{Hess} r\| \leq \frac{s'_{K_{\min}}(r)}{s_{K_{\min}}(r)},$$

从而 $\|\dot{J}(s)\| \leq \|J\|$|海森矩阵 $r\| \leq \frac{s'_{K_{\min}}(s)}{s_{K_{\min}}(s)} s_{K_{\min}}(s) \leq s'_{K_{\min}}(s).$

此外，Lezcano-Casado (2020，定理 4.11) 推出（将 w 替换为 rw）

$$K(s) \leq \frac{16}{3}s_{K_{\min}}(s/2)^2 L_R s_{K_{\min}}(s).$$

因此，我们有

$$\|\dot{K}(t)\| \leq \int_0^t L_R(\frac{16}{3}s_{K_{\min}}(s/2)^2 L_R s_{K_{\min}}(s)) + 2(L_R s_{K_{\min}}(2s))ds.$$

注意，在我们的情形中，$s_{K_{\min}}(t) = \sinh(\sqrt{|K_{\min}|}t)\sqrt{|K_{\min}|}$，因此我们有 $s_{K_{\min}}(s/2) \leq s_{K_{\min}}(s)$。

对于第一项，我们有

$$\begin{aligned}\int_0^t L_R(\frac{16}{3}s_{K_{\min}}(s/2)^2 L_R s_{K_{\min}}(s))ds &\leq L_R^2\frac{16}{3}\int_0^t s_{K_{\min}}(s)^3 ds \leq L_R^2\frac{16}{3}\int_0^t s_{K_{\min}}(t)^3 ds \\ &= L_R^2\frac{16}{3}ts_{K_{\min}}(t)^3.\end{aligned}$$

对于第二项，

$$\int_0^t 2(L_R s_{K_{\min}}(2s))ds \leq 2L_R s_{K_{\min}}(t)^2.$$

综合起来，我们得到

$$\|\dot{K}(t)\| \leq L_R^2\frac{16}{3}ts_{K_{\min}}(t)^3 + 2L_R s_{K_{\min}}(t)^2.$$

□

E.5 涉及黎曼对数映射的界

这里，我们给出了 $\|(d\text{对数映射}_x)_y\|$、 $\left\|\left(\nabla(d\text{对数映射}_x)\right)_y\right\|$、 $\|\nabla_x \text{对数映射}_x(y)\|_{\mathrm{op}}$、 $\|\nabla_x \text{对数映射}_x(y)\|_{\mathrm{op}}^2$、 |散度$_x$对数映射$_x(y)$|以及 |散度$_x$对数映射$_x(y)|^2$的界。当流形为对称正定矩阵（$n$）时，我们还给出了 ‖梯度散度$_x$对数映射$_x(y)$‖ 的一个界。本节中，对于不同的微分算子 d和 ∇，请注意， ∇对数映射$_x(x_1)$是协变导数（将对数映射$_x(x_1)$视为向量场，并固定 x_1）。另一方面， d对数映射$_{x_1}$表示固定基点的情形。

在给出结果之前，我们还回顾一些事实。设 M 是维数为 d的 Hadamard 流形。固定 $x \in M$，并令 $y \in M$、 $y \neq x$。那么：

(i) $\|\operatorname{Log}_x(y)\| = d(x,y).$

(ii) $\operatorname{grad}_x\left(\frac{1}{2}d(x,y)^2\right) = -\operatorname{Log}_x(y).$

(iii) $\operatorname{div}_x \operatorname{Log}_x(y) = -\Delta_x\left(\frac{1}{2}d(x,y)^2\right).$

引理 E.12 (d对数映射$_x$和 $\nabla(d$对数映射$_x)$ **的界).** 设 M是维数为d的 ***Hadamard*** 流形。固定 $x \in M$。那么，我们有

$$\begin{aligned}\|(d\operatorname{Log}_x)_y\| &\leq 1,\\ \left\|\left(\nabla(d\operatorname{Log}_x)\right)_y\right\| &\leq \left\|\left(\nabla(d\operatorname{Exp}_x)\right)_{\operatorname{Log}_x(y)}\right\|.\end{aligned}$$

证明。 在 Hadamard 流形上，$\operatorname{Exp}_x\colon T_xM \to M$是全局微分同胚，因此 $\operatorname{Log}_x = \operatorname{Exp}_x^{-1}$在 M上光滑。令 $u = \operatorname{Log}_x(y)$。则 $\operatorname{Exp}_x(u) = y$，由反函数定理可得

$$(d\operatorname{Log}_x)_y = \left((d\operatorname{Exp}_x)_u\right)^{-1} = \left((d\operatorname{Exp}_x)_{\operatorname{Log}_x(y)}\right)^{-1}.$$

对等式两边取范数，并回忆 $\left\|\left((d\operatorname{Exp}_x)_{\operatorname{Log}_x(y)}\right)^{-1}\right\| \leq 1$，即可得到第一个界。

现在我们对该恒等式求导

$$(d\operatorname{Exp}_x)_{\operatorname{Log}_x(\cdot)} \circ (d\operatorname{Log}_x)_{\cdot} = Id_{T\cdot M}$$

在 y 处沿 $w_1 \in T_yM$ 方向取协变导数，然后将所得算子作用于 $w_2 \in T_yM$。利用莱布尼茨法则以及 $\nabla(Id) = 0$，可得

$$\left(\nabla(d\operatorname{Exp}_x)\right)_{\operatorname{Log}_x(y)}\left[(d\operatorname{Log}_x)_y w_1,\, (d\operatorname{Log}_x)_y w_2\right] + (d\operatorname{Exp}_x)_{\operatorname{Log}_x(y)}\left(\left(\nabla(d\operatorname{Log}_x)\right)_y[w_1,w_2]\right) = 0.$$

重新排列各项，可得

$$\begin{aligned}&\left(\nabla(d\operatorname{Log}_x)\right)_y[w_1,w_2]\\ &= -\left((d\operatorname{Exp}_x)_{\operatorname{Log}_x(y)}\right)^{-1}\left(\left(\nabla(d\operatorname{Exp}_x)\right)_{\operatorname{Log}_x(y)}\left[\left((d\operatorname{Exp}_x)_{\operatorname{Log}_x(y)}\right)^{-1}w_1,\, \left((d\operatorname{Exp}_x)_{\operatorname{Log}_x(y)}\right)^{-1}w_2\right]\right).\end{aligned}$$

取范数，并回顾 $\|\left((d\text{指数映射}_x)_{\operatorname{Log}_x(y)}\right)^{-1}\| \leq 1$，我们得到

$$\left\|\left(\nabla(d\operatorname{Log}_x)\right)_y\right\| \leq \left\|\left(\nabla(d\operatorname{Exp}_x)\right)_{\operatorname{Log}_x(y)}\right\|.$$

□

引理 E.13（关于 $\|\nabla_x$ 对数映射$_x(y)\|$的界）。设 M 为维数为 d 的 *Hadamard* 流形，其截面曲率下界为 $K_{\min} \le 0$。固定 $x \in M$ 与 $y \in M$、$y \neq x$。则

$$\|\nabla_x \operatorname{Log}_x(y)\|_{\mathrm{op}} \le 1 + d(x,y)\,\frac{s'_{K_{\min}}(d(x,y))}{s_{K_{\min}}(d(x,y))},$$
$$\|\nabla_x \operatorname{Log}_x(y)\|_{\mathrm{op}}^2 \le 2 + 2\,d(x,y)^2\Big(\frac{s'_{K_{\min}}(d(x,y))}{s_{K_{\min}}(d(x,y))}\Big)^2.$$

证明。令 $r(\cdot) := d(\cdot,\ y)$，则 $r(x) = d(x,\ y)$。由于 r 梯度 $_x r = -$对数映射 $_x(y)$（因为梯度 $_x(\frac{1}{2}r^2)$ $= r$ 梯度 $_x r$），我们有

$$\operatorname{Log}_x(y) = -\,r(x)\ \operatorname{grad}_x r(x).$$

对 x作协变微分，得到

$$\nabla_x \text{对数映射}_x = -\nabla_x\big(r(x)\ \operatorname{grad}_x r(x)\big) = -\big(\nabla_x r(x)\big) \otimes \operatorname{grad}_x r(x)\ -\ r(x)\,\nabla_x \operatorname{grad}_x r(x)$$

取算子范数并利用 $|$梯度 $_x r(x)\| = 1$，得到

$$\|\nabla_x \operatorname{Log}_x(y)\|_{\mathrm{op}} \le \|\operatorname{grad}_x r(x)\|^2 + r(x)\,\|\nabla \operatorname{grad}_x r(x)\|_{\mathrm{op}} = 1 + d(x,y)\,\|\nabla \operatorname{grad}_x r(x)\|_{\mathrm{op}}.$$

根据截面曲率 $\ge K_{\min}$ 下的 Hessian 比较定理，我们有

$$\|\nabla \operatorname{grad}_x r(x)\|_{\mathrm{op}} \le \frac{s'_{K_{\min}}(d(x,y))}{s_{K_{\min}}(d(x,y))}.$$

结合以上估计可得第一个不等式。平方形式的界由 $(a+b)^2 \le 2a^2 + 2b^2$ 得到，其中 $a = 1$ 且 $b = d(x,\ y)\, s'_{K_{\min}}(d(x,\ y))\, s_{K_{\min}}(d(x,\ y))$。 □

引理 E.14 (散度的界 $_x$ 对数映射 $_x(y)$ 和 $|$散度 $_x$ 对数映射 $_x(y)|^2$)。设 M 是一个维数为 d、截面曲率下界为 $K_{\min} \le 0$的 *Hadamard* 流形。固定 $x \in M$ 和 $y \in M,\ y \neq x$。则

$$|\operatorname{div}_x \operatorname{Log}_x(y)| \le 1 + (d-1)\,d(x,y)\,\frac{s'_{K_{\min}}(d(x,y))}{s_{K_{\min}}(d(x,y))},$$
$$|\operatorname{div}_x \operatorname{Log}_x(y)|^2 \le 2 + 2(d-1)^2\,d(x,y)^2\Big(\frac{s'_{K_{\min}}(d(x,y))}{s_{K_{\min}}(d(x,y))}\Big)^2.$$

证明。设 $r(\cdot) := d(\cdot, y)$。于是有 div $_x$对数映射 $_x(y) = -\Delta_x\Big(\frac{1}{2}r(x)^2\Big)$。利用拉普拉斯的乘积法则，可得 $\Delta(\frac{1}{2}r^2) = \langle$梯度 r，梯度 $r\rangle + r\Delta r = 1 + r\Delta r$。对于上界，由 Sec $\ge K_{\min}$ 条件下的拉普拉斯比较定理可得

$$\Delta r \le (d-1)\,\frac{s'_{K_{\min}}(d(x,y))}{s_{K_{\min}}(d(x,y))}.$$

代入 $|$散度 $_x$对数映射 $_x(y)| = \Delta_x(\frac{1}{2}r^2)$，可得

$$|\operatorname{div}_x \operatorname{Log}_x(y)| \le 1 + (d-1)\,d(x,y)\,\frac{s'_{K_{\min}}(d(x,y))}{s_{K_{\min}}(d(x,y))}.$$

最后，利用 $(a+b)^2 \le 2a^2 + 2b^2$ 以及 $a = 1$ 和 $b = (d-1)\,d(x,\ y)\, s'_{K_{\min}}(d(x,\ y))\, s_{K_{\min}}(d(x,\ y))$ 即可得到所需结果。 □

此外，在对称正定矩阵 (n) 上，我们有以下结果。

引理 E.15 (|| 梯度 散度 对数映射 || 在对称正定矩阵 (n))。 设 对称正定矩阵 (n) 为实对称正定矩阵构成的流形，并赋予仿射不变黎曼度量

$$\langle U, V\rangle_P := \operatorname{Tr}\left(P^{-1}UP^{-1}V\right), \qquad U, V \in T_P \operatorname{SPD}(n) = Sym(n).$$

固定 $y \in$ 对称正定矩阵 (n) 并定义

$$r(x) := d(x, y), \qquad f(x) := d(x, y)^2 = r(x)^2, \qquad d := \dim(\operatorname{SPD}(n)) = \frac{n(n+1)}{2}.$$

于是我们有

$$\| \operatorname{grad} \operatorname{div} \operatorname{Log} \| \leq \frac{\sqrt{2}d}{2}\left(2\left(1 + r\frac{s'_{K_{\min}}(r)}{s_{K_{\min}}(r)}\right)\right)^{\frac{3}{2}}.$$

证明。 我们首先证明，在仿射不变度量下，对称正定矩阵 (n) 是 $PD(n, \mathbb{C})$ 的一个全测地子流形。令 PD$(n, \mathbb{C})$ 表示配备相同仿射不变度量 $\langle U, V\rangle_P = \operatorname{Tr}(P^{-1}UP^{-1}V)$ 的复厄米正定矩阵流形。定义一个光滑等距映射如下:

$$\sigma : \operatorname{PD}(n, \mathbb{C}) \to \operatorname{PD}(n, \mathbb{C}), \qquad \sigma(P) := \overline{P}.$$

No注意到 σ 是一个等距映射，因为对于任意 $P \in$ PD $(n, \mathbb{C})$ 和 $U, V \in Herm(n)$,

$$\langle d\sigma_P[U], d\sigma_P[V]\rangle_{\sigma(P)} = \operatorname{Tr}\left(\overline{P}^{-1}\,\overline{U}\,\overline{P}^{-1}\,\overline{V}\right) = \overline{\operatorname{Tr}\left(P^{-1}UP^{-1}V\right)} = \operatorname{Tr}\left(P^{-1}UP^{-1}V\right) = \langle U, V\rangle_P,$$

其中我们用到了，对于厄米矩阵 $P^{-1}UP^{-1}V$， $\in \mathbb{R}$， P, U, V。因此我们看到， σ 的不动点集恰好是对称正定矩阵 (n)。根据 Kobayashi (1972，定理 5.1)，等距映射的不动点集是一个（嵌入的）全测地子流形。因此，对称正定矩阵 (n) 是 PD $(n, \mathbb{C})$ 中的全测地子流形。

令 $F(P) := d_{\operatorname{PD}(n,\mathbb{C})}(P, y)^2$ 表示到 $y \in$ 对称正定矩阵 $(n) \subset PD(n, \mathbb{C})$ 的距离平方，将其视为 $PD(n, \mathbb{C})$ 上的函数。根据 Hirai 等人（2023，定理 1.4）， F 在 PD $(n, \mathbb{C})$ 上是 2-自协调的：对于任意 $P \in$ PD $(n, \mathbb{C})$ 以及 $u, v, w \in T_P$PD $(n, \mathbb{C})$，均有

$$|(\nabla^3 F)_P(u, v, w)| \leq \sqrt{2}\sqrt{(\nabla^2 F)_P(u, u)}\sqrt{(\nabla^2 F)_P(v, v)}\sqrt{(\nabla^2 F)_P(w, w)}.$$

现在将 F 限制在全测地子流形对称正定矩阵 $(n$上。根据定义，对于全测地子流形，第二基本形式为零，而高斯公式 (Lee，2018，定理 8.2) 表明， ∇ 在对称正定矩阵 $(n$上的联络由 PD $(n,\ \mathbb{C})$ 上的联络诱导。此外，对称正定矩阵 $(n$上的测地距离函数由 PD $(n,\ \mathbb{C}$上的测地距离函数诱导，因此距离平方的自协调性在对称正定矩阵 $(n$上同样成立。

现在我们利用自协调性来界定 || 梯度 散度 对数映射 ||。令 $f(x) = d_{\operatorname{SPD}}(x, y)^2$，并注意在 Hadamard 流形 对称正定矩阵$(n)$ 上,

$$\| \operatorname{grad}_x \operatorname{div}_x \operatorname{Log}_x(y)\| = \tfrac{1}{2}\| \operatorname{grad}_x \Delta_x f(x)\|.$$

令 $\{e_i\}_{i=1}^d$ 为 T_x 对称正定矩阵 $(n$的一组正交归一基。该正交归一基可以扩展到 x的一个邻域，从而得到一个正交规范标架： $\nabla e_i|_x = 0, \forall i$ (Lee，2018，习题 5-21)。因此，我们可以如下对 Δf 求导。对于任意单位 $w \in T_x$ 对称正定矩阵 (n),

$$\langle \operatorname{grad} \Delta f, w\rangle = \sum_{i=1}^{d}(\nabla^3 f)(w, e_i, e_i) + 2\nabla^2 f(\nabla_w e_i, e_i) = \sum_{i=1}^{d}(\nabla^3 f)(w, e_i, e_i).$$

根据自协调性，我们有 $|(\nabla^3 f)(w,\ e_i,\ e_i)| \le \sqrt{2}\,\sqrt{(\nabla^2 f)(w,\ w)\,(\nabla^2 f)(e_i,\ e_i)}$。因此

$$\|\operatorname{grad}\operatorname{div}\operatorname{Log}\| \le \frac{\sqrt{2}d}{2}\sqrt{\|\nabla^2 f\|_{\mathrm{op}}^3}.$$

记 $f = r^2$，我们有

$$\nabla^2 f = 2\,dr\otimes dr + 2r\,\nabla^2 r, \qquad \Delta f = 2 + 2r\,\Delta r,$$

利用 $\|$梯度 $r\| = 1$。在截面曲率 $\ge K_{\min}$ 的条件下，Hessian 比较定理给出 $\|\nabla^2 r\|_{\mathrm{op}} \le {}^{s'_{K_{\min}}(r)}\,{}_{s_{K_{\min}}(r)}$，因此

$$\|\nabla^2 f\|_{\mathrm{op}} \le 2 + 2r\,\|\nabla^2 r\|_{\mathrm{op}} \le 2\Big(1 + r\,\frac{s'_{K_{\min}}(r)}{s_{K_{\min}}(r)}\Big).$$

将其代入前一个估计可得

$$\|\operatorname{grad}\operatorname{div}\operatorname{Log}\| \le \frac{\sqrt{2}d}{2}\left(2\Big(1 + r\,\frac{s'_{K_{\min}}(r)}{s_{K_{\min}}(r)}\Big)\right)^{1.5}.$$

□

E.6 辅助结果

给定一个矩阵 $A(x)$，v 中的方向导数可通过以下方式计算

$$D\log\det(A(x))[v] = \operatorname{tr}(A^{-1}DA(x)[v]).$$

因此

$$\|\nabla\log\det(A(x))\|_{\mathrm{op}} = \sup_{\|v\|=1} |\operatorname{tr}(A^{-1}DA(x)[v])| \le d\|A^{-1}\| \times \|DA(x)\|_{\mathrm{op}}.$$

此外，利用乘积法则，并注意到 $DA^{-1}(x) = -A^{-1}(x)DA(x)A^{-1}(x)$，我们得到

$$\begin{aligned}
&D^2\log\det(A(x))[v,v]\\
=&\operatorname{tr}(DA^{-1}(x)[v]DA(x)[v]) + \operatorname{tr}(A^{-1}(x)D^2A(x)[v,v])\\
=&-\operatorname{tr}(A^{-1}(x)DA(x)[v]A^{-1}(x)DA(x)[v]) + \operatorname{tr}(A^{-1}(x)D^2A(x)[v,v]).
\end{aligned}$$

因此，我们有

$$\|\nabla^2\log\det(A(x))\|_{op} \le d\|A^{-1}(x)\|^2 \times \|\nabla A(x)\|^2 + d\|A^{-1}(x)\| \times \|\nabla^2 A(x)\|.$$

这可以总结为以下结果，并用于引理 E.8。

引理 E.16 (行列式对数的导数)。设 $A(x)$ 是内积空间上可逆线性映射的光滑族，该空间的维数为d。那么，对任意单位向量 v,

$$D\log\det(A(x))[v] = \operatorname{tr}\left(A(x)^{-1}DA(x)[v]\right).$$

因此,

$$\|\nabla\log\det(A(x))\| = \sup_{\|v\|=1}\left|\operatorname{tr}\left(A(x)^{-1}DA(x)[v]\right)\right| \le d\,\|A(x)^{-1}\|\,\|\nabla A(x)\|, \tag{33}$$

$$\|\nabla^2\log\det(A(x))\|_{\mathrm{op}} \le d\,\|A(x)^{-1}\|^2\,\|\nabla A(x)\|^2 + d\,\|A(x)^{-1}\|\,\|\nabla^2 A(x)\|. \tag{34}$$