arXiv:2305.03215v1 [math.ST] 5 May 2023

度量空间中的测地凸 M 估计

Victor-Emmanuel Brunel *

摘要：我们研究非线性空间上测地凸 M 估计的渐近性质。具体而言，我们证明，除代价函数的测地凸性外，只需非常弱的假设，便可得到一致性和渐近正态性；这两者是统计推断中的基本性质。我们的结果扩展了凸 M 估计的欧几里得理论；此外，我们还推广了非线性空间上的极限定理。这些定理本质上此前仅适用于重心，而我们的结果使人们能够考虑通过非光滑 M 估计过程定义的稳健替代方法。

关键词与短语： M 估计、度量空间、黎曼流形、CAT 空间、测地凸性、重心、稳健位置估计。

1. 预备知识

1.1 引言

M估计，或经验风险最小化，是统计学和机器学习中普遍存在的问题。在许多问题中，人们试图最小化一个具有期望形式的函数，即 $\Phi(x) = \mathbb{E}[\phi(Z,x)]$, 其中， Z 是一个分布未知的随机变量，x 是目标变量，它存在于某个目标空间中，而 ϕ 可以看作一个代价函数。分类、回归、位置估计和最大似然估计，都是此类问题的标准实例。在实践中， Z 的分布是未知的，因此无法计算函数 Φ 。所以，人们通常最小化形如 $\hat{\Phi}_n(x) = n^{-1}\sum_{i=1}^n \phi(Z_i,x)$ 的代理函数，其中 Z_1，……， Z_n 是 Z 的可用独立同分布副本。这种方法称为 M估计，或经验风险最小化。当目标空间是欧氏空间时，关于$\hat{\Phi}_n$的极小值点$\hat{x}_n$作为 Φ的极小值点 x^* 的估计量所具有的保证，已有大量研究结果。渐近结果（如一致性和渐近正态性）是经典且广为人知的，尤其是在对代价函数 ϕ [53]作出技术性光滑性假设时；当代价函数关于目标变量 [28, 44]为凸函数时，其中大多数假设都可以舍弃。

在本研究中，我们关注的是这样一种框架：目标空间不是欧几里得空间，甚至不具有任何线性结构。当然，该空间至少需要配备一个度量，以便衡量某种估计程序的准确性。因此，我们考虑一个度量空间作为目标空间，并将其记为 (M,d)。

统计学和机器学习越来越多地需要处理位于非线性空间中的数据：空间统计学（例如方向数据）、计算机断层成像（例如形状统计学中采集的商空间数据，此类数据仅在刚性变换意义下确定）、经济学（例如最优传输，其中数据是测度）等。此外，以……形式编码的数据和/或目标参数

*CREST-ENSAE，victor.emmanuel.brunel@ensae.fr

为超高维向量的数据可能具有小得多的内在维数：例如，这些向量位于环境欧几里得空间中的低维子流形上。在这种情况下，利用所研究问题可能具有的非线性几何结构，可以有力地显著降低其维数。这正是流形学习 [1,26] 或生成对抗网络 [50]背后的核心动机。尽管越来越多的算法被开发出来，用于在非线性空间中进行计算，尤其是优化 [7,19,42,46,55,56],但从统计学角度来看，我们对此仍知之甚少。

度量空间上 M 估计的一个受到广泛关注的具体实例是重心。给定 n 个点 $x_1, \ldots, x_n \in M$，重心是平方距离之和 $\sum_{i=1}^n d(x,x_i)^2$ 的任意极小值点 $x \in M$。容易验证，若 (M,d) 是欧几里得空间或希尔伯特空间，则极小值点唯一，且等于 $x_1, \ldots, x_n$ 的平均值。重心将平均值的概念推广到了缺乏线性结构的空间。更一般地，给定 M 上的概率测度 P，只要对至少一个（因而也对所有）$x \in M$ 值而言积分有限，就可以将 P 的重心定义为 $\int_M d(x,z)^2 \mathrm{d}P(z)$ 的任意极小值点 $x \in M$。20 世纪 40 年代， [24] 最初将重心引入统计学；后来 [33]，也对其进行了研究，并将其称为 Fr´echet 均值或卡尔歇均值。重心在形状统计学 [34]、最优传输 [2, 5, 6, 18, 20, 36–38] 以及矩阵分析 [8–10] 领域得到了推广。对于各种设定下的经验重心，其渐近理论已有相当深入的研究，尤其包括大数定律[51, 57] ，以及黎曼流形上的中心极限定理（为了推导中心极限定理，自然需要在 M 上施加光滑结构）[11–13, 22, 23, 32, 35]。目前已有一些有限样本保证，但重心的非渐近统计理论仍相当有限 [3,25,39,49,51]。对于黎曼流形上更一般的 M 估计量，例如 p 均值，也已有渐近定理，其中 $\mathsf{Z} = X$ 且 $\phi(z,x) = d(x,z)^p$，对某个 $p \geq 1$ [49] 成立；几何中位数是其中一个特例（$p = 1$），但目前已知的速率都不是最优的。特别是，除重心之外，一般情形下标准的 $n^{-1/2}$ 速率尚不清楚。值得注意的是，黎曼流形上关于重心的所有中心极限定理都依赖代价函数 $\phi = d^2$ 的光滑性，并通过在局部坐标图中进行泰勒展开，采用光滑 M 估计的标准技术 [53] 。此外，这些定理还假设数据几乎必然落在一个半径较小的球内；这一条件基本上保证了平方距离函数 d^2 关于第二个变量的凸性，尽管这项综合性质并未在这一研究方向中得到利用。例如，采用这些技术，无法推导出几何中位数或其他重心稳健替代方法的渐近正态性。

1.2 贡献

在本文中，我们在非常温和的假设下证明了测地凸 M-估计量的强相合性和渐近正态性。我们重新得到了针对直径较小的黎曼流形上重心所证明的中心极限定理，但研究范围涵盖更广泛的位置估计量，包括重心的稳健替代方法。与欧几里得情形一样，我们的结果表明凸性是一项强大的性质，因为只需非常少的假设，它就能带来强有力的学习保证。

我们的所有结果都是渐近的，但值得一提的是，在非线性空间中证明学习保证是一项极具挑战性且仍在进行中的研究课题，因为这需要度量几何和微分几何中的非标准工具。特别是，即使对于重心这类简单的估计量，可获得的非渐近学习保证也非常少，而在该框架下，渐近理论仍然是一个活跃的研究领域（例如，理解非标准渐近速率，即所谓的黏滞与涂抹中心极限定理 [23, 30]）。此外，渐近正态性等渐近保证还能提供基准，这些基准对于有限样本理论同样十分重要。

2. 背景与符号

2.1 测地空间

设 (M,d) 是一个满足以下性质的完备可分度量空间：对于所有 $x,y\in M$，存在一个连续函数 $\gamma:[0,1]\to M$，使得 $\gamma(0)=x$、$\gamma(1)=y$，并且对于所有 $s,t\in[0,1]$，满足 $d(\gamma(s),\gamma(t))=|s-t|d(x,y)$。这样的函数 γ 称为从 x 到 y 的单位速测地线，应将其理解为从 x 到 y 的一条（不一定唯一的）最短路径。我们用 $\Gamma_{x,y}$ 表示所有此类函数的集合。空间 (M,d) 随后称为测地空间。测地空间的例子包括欧几里得空间、欧几里得球面、度量图、瓦瑟斯坦空间、商空间等。[15,16]。测地空间为凸性的概念提供了一个自然的推广。

Definition 1. 若对所有 x、$y\in M$、$\gamma\in\Gamma_{x,y}$ 以及 $t\in[0,1]$、$f(\gamma(t))\le(1-t)f(x)+tf(y)$，函数 $f:M\to\mathbb{R}$ 就称为测地凸（简称凸）。

本文进一步假设（M,d）是正则空间，即所有有界闭集都是紧集。这一假设在实际应用中或许显得有所限制，但从较高层次来看，它实际上仅排除了无限维空间。事实上，霍普夫-里诺定理 [15, 第 1 章，命题 3.7]指出，只要（M,d）是完备且局部紧的，它就是正则空间。例如，黎曼流形的霍普夫-里诺定理指出，任何完备的（有限维）黎曼流形都是正则测地度量空间 [21, 第 7 章，定理 2.8]。

2.2 黎曼流形

黎曼流形 (M,g) 是一个有限维光滑流形 M，其配备了黎曼度量 g，即切空间上的标量积的光滑族。关于微分几何和黎曼流形的清晰介绍，请参阅 [21] 和 [40,41]。由黎曼度量 g 在 M 上诱导的距离记为 d。在本文中，为简单起见，我们仅考虑无边界的黎曼流形，尽管我们的结果很容易推广到有边界的情形。

M的切丛记作 TM，即$TM=\bigcup T_xM$其中，对于所有 $x\in M$、T_xM，$x\in M$是 M 在 x处的切空间。对于所有 $x\in M$，我们还分别令 $\mathrm{Exp}_x:T_xM\to M$ 和 $\mathrm{Log}_x:M\to T_xM$ 表示 x处的指数映射和对数映射。后者可能未定义于整个切空间上，但至少总是在原点的某个邻域内有定义。为便于理解，在给定点 $x\in M$处，对于给定向量 $u\in T_xM$，$\mathrm{Exp}_x(u)$是欧几里得空间中“$x+u$”的黎曼类比：从点 x出发，加上一个向量 u，其方向表示移动的方向，而其范数（由黎曼度量 g给出）表示移动的距离，从而到达记作 $\mathrm{Exp}_x(u)$的点。反过来，给定 x、$y\in M$，$\mathrm{Log}_x(y)$（当其有明确定义时）是“$y-x$”的黎曼类比，也就是说，从 x出发，T_xM中的哪个向量促成了移动到 y。

对于所有 $x\in M$，我们用 $\cdot,\cdot_x$表示 T_xM 上由 g 继承的标量积，并用 $\|\cdot\|_x$ 表示相应的范数。

令 $x,y\in M$和 $\gamma\in\Gamma_{x,y}$为从 x 到 y的一条测地线（按第 2.1 节中的定义）。若 x 和 y 足够接近，则根据命题 6.11]，γ 是唯一的 [41,。在这种情况下，对所有 $t\in[0,1]$,$\gamma(t)=\mathrm{Exp}_x(t\dot\gamma(0))$，其中 $\dot\gamma(0)\in T_xM$ 是 γ 在 $t=0$ 处的导数，由 $\dot\gamma(0)=\mathrm{Log}_x(y)$给出。我们将沿 γ从 x 到 y 的平行移动映射记为 $\pi_{x,y}:T_xM\to T_yM$，不特别注明其对 γ的依赖（若 y 与 x足够接近，则该依赖是明确无歧义的）。映射 $\pi_{x,y}$ 是一个线性等距映射，即对所有 u，$v\in T_xM$，均有 $u,v_x=\pi_{x,y}(u)\pi_{x,y}(v)y$。此外，还有 $\pi_{x,y}(\dot\gamma(0))=-\dot\gamma(1)$。直观地说，

平行移动是沿流形上的给定路径比较不同切空间中的向量的一种重要工具。

有关本研究数学推导中所用黎曼流形的更多细节和性质，请读者参阅附录B。

2.3 M-估计

设 Z 是某个抽象空间，配备一个 σ-代数 F 和一个概率测度 P。令$\phi : \mathsf{Z} \times M \to \mathbb{R}$ 满足以下条件：

(i) 对所有 $x \in M$，映射 $\phi(\cdot, x)$ 关于 P 可测且可积；(ii) 对 P-几乎所有 $z \in \mathsf{Z}$， $\phi(z, \cdot)$ 是凸的。

最后，设 $Z, Z_1, Z_2, \ldots$ 是取值于 Z、服从分布 P的独立同分布随机变量，并定义函数

$$\Phi(x) = \mathbb{E}[\phi(Z, x)] \quad \text{and} \quad \hat{\Phi}_n(x) = n^{-1} \sum_{i=1}^{n} \phi(Z_i, x),$$

对所有 $x \in M$ 以及所有正整数 n均成立。最后，我们用 M^* 表示 Φ的极小化点集合。注意，由于 Φ的凸性，极小化集合 M^* 是凸集，即对所有 x、 $y \in M^*$ 以及所有$\gamma \in \Gamma_{x,\, y}$、 $\gamma([0,1]) \subseteq M^*$。

重要的例子包括 $\mathsf{Z} = M$ 且 ϕ 是度量的函数的情形，这将得到位置估计。具体而言，假设 $\mathsf{Z} = M$且对所有 z、 $x \in M$，有 $\phi(z, x) = \ell(d(z, x))$，其中 $\ell : [0, \infty) \to [0, \infty)$ 是一个非递减凸函数。那么，确保函数 $\phi(z, \cdot)$ 凸的一个自然框架是具有曲率上界的空间，也称为 CAT 空间。这里，我们仅对 CAT 空间给出直观定义。关于 CAT 空间的更详细介绍，请参阅附录 D。

取任意实数 κ。如果 M 中所有足够小的三角形都比常曲率为 κ 的空间中的相应三角形更薄，即当 $\kappa > 0$ 时为球面，当 $\kappa = 0$ 时为欧几里得空间，当 $\kappa < 0$ 时为双曲空间，那么我们称 (M, d) 是 CAT(κ)（或称其具有全局曲率上界 κ）。这种设定非常自然，可用于理解点到该点的距离函数的凸性。取$x_0 \in M$。函数 $d(\cdot, x_0)$ 是凸的，当且仅当对所有 y、 $z \in M$ 以及所有 $\gamma \in \Gamma_{y,\, z}$， $d(\gamma(t)\ x_0) \leq (1-t)d(y, x_0) + td(z, x_0)$，这控制了三角形中从 x_0 到对边上任意一点的距离，其中该三角形的顶点为 x_0、 y、 z，而 x_0 是其中一个顶点。换句话说，它说明顶点为x_0、 y、 z 的三角形必须有多薄。

因此，如果 $\mathsf{Z} = M$ 是 CAT(κ) 空间且直径至多为 D_κ （定义见附录 D）， [45,命题 3.1] 保证，对于所有 $z \in M$， $d(z, \cdot)$ 在 M 上是凸的；因此，只要 $\ell : [0, \infty) \to [0, \infty)$ 本身是非递减凸函数，$\phi(z, \cdot) = \ell(\phi(z, \cdot))$ 也为凸函数。在这种情况下，函数 ℓ 的重要例子包括：

(i) $\ell(u) = u^2$：那么， Φ 的极小值点是 P的重心，而 $\hat{} \Phi_n$ 的极小值点是 Z_1的经验重心，$\ldots \ldots Z_n$(ii) $\ell(u) = u$：那么， Φ 的极小值点是 P的几何中位数。(iii) 更一般地，如果 $\ell(u) = u^p$，$p \geq 1$，则 Φ 的极小值点称为 P的 p-均值。(iv) $\ell(u) = u^2$ （当 $0 \leq u \leq c$ 时）， $\ell(u) = c(2u - c)$ （当 $u > c$ 时），其中 $c > 0$ 是一个固定参数。Huber [31] 引入了该函数，以构造均值的低偏差稳健估计量：它在重心估计与几何中位数估计之间提供了一种插值。

CAT 空间的重要例子包括欧几里得空间、球面、具有有界截面曲率的单连通黎曼流形、度量树等。在附录 D中，我们提供了更为详尽的例子列表。

3. 一致性

本节介绍关于测地凸 M-估计量一致性的首批主要结果。在这里，我们在两种情形下开展研究。首先，我们假设 $\phi(z,\cdot)$ 是凸的，并且满足 Lipschitz 条件，其常数不依赖于 $z \in Z$。尽管这一假设涵盖了上述位置估计中提到的大多数情形，但它仍具有一定限制性。在第二种情形下，我们只假设其具有凸性，但考虑 M 为黎曼流形的情况。随后，我们将讨论研究结果可能的扩展。

3.1 Lipschitz 情形

这里，我们假设存在 $L > 0$，使得对于所有 $z \in Z$，$\phi(z,\cdot)$ 在 M 上是 L-Lipschitz 的。例如，如果 M 的直径有界，且 $\phi = \ell \bigcirc d$，其中 $\ell : [0,\infty$ 是某个局部利普希茨函数（例如上文提到的位置估计函数），则满足这一条件。

回顾一下，我们用 M^* 表示 Φ 的极小化点集合。以下定理并不要求 Φ 有唯一的极小值点：它断言，Φ_n 的任意极小值点最终都会以概率 1 任意接近 M^*。

THEOREM 1. 设 M 是一个完备测地度量空间。假设 M^* 非空且有界。对于 $n \geq 1$，令 $\hat{x}_n$ 为 $\hat{\Phi}_n$ 的一个极小值点。则 $d(\hat{x}_n, M^*) \xrightarrow[n\to\infty]{} 0$ 几乎必然成立。

简单调整证明即可表明，如果 $\hat{x}_n$ 并非恰好是 Φ_n 的极小值点，而是满足 $\Phi_n(\hat{x}_n) \leq \min_{x\in M} \Phi_n(x) + \varepsilon_n$，其中 ε_n 是任意趋于零的非负误差项，且 $n \to \infty$ 几乎必然成立，则同样的结论仍然适用。

定理 1 的证明基于 [47, 定理 108]的以下推广。证明是对 [47,定理 108]证明的直接调整。

LEMMA 1. 设 $L > 0$ 和 $(f_n)_{n\geq 1}$ 是 M上的一列 L-Lipschitz 凸函数。假设存在一个函数 $f : M \to \mathbb{R}$ 以及 M 的一个稠密子集 M_0，使得对于所有$x \in M_0$，$f_n(x) \xrightarrow[n\to\infty]{} f(x)$。则，$f$ 是凸的、L-Lipschitz 的，并且 f_n 在 M的任意紧子集上一致收敛于 f。

这个引理解释了为何本节要求 $\phi(z,\cdot)$ 满足 Lipschitz 条件，且其常数不依赖于 $z \in Z$。如果 M 是欧几里得空间，那么引理 1 中的 Lipschitz 假设便无须成立，因为在任意紧子集上，它都可由函数 $f_n, n \geq 1$, 的凸性和逐点收敛推出，参见 [47, 定理 10。6]然而，据我们所知，在此处讨论的更一般情形下，这一结论未必成立（这或许值得进一步研究）。

作为引理 1的推论，我们现在陈述以下结果；它们对定理 1的证明至关重要。证明见附录C。

LEMMA 2. 设 $L > 0$ 和 $(F_n)_{n\geq 1}$ 是定义在 M上的一列实值、L-Lipschitz 凸随机函数。设 F是定义在 M上的一个（可能为随机的）函数，并假设对所有 $x \in M$，$F_n(x)$ 几乎必然（分别地，依概率）收敛于 $F(x)$。那么，该收敛在 M的任意紧子集 K 上一致成立，即 $\sup_{x\in K} |F_n(x) - F(x)|$ 几乎必然（分别地，依概率）收敛于零。

PROOF OF THEOREM 1。设 $\Phi^* = \min_{x\in M}\Phi(x)$ 为 M 上 Φ 的最小值，并固定任意一个 $x^* \in M^*$。固定 $\varepsilon > 0$，并令 $K_\varepsilon = \{x \in M : d(x, M^*) = \varepsilon\}$。由于 M^* 有界，K_ε 有界。此外，由于距离是连续的，K_ε 是闭集。因此，由于我们假设 (M, d) 是完备的，它是紧的。由于 Φ 是 Lipschitz 的，因此连续，从而有 $\eta := \min_{x\in K_\varepsilon}\Phi(x) - \Phi^* > 0$。此外，根据大数定律，对所有 $x \in M$，$\Phi_n(x) \xrightarrow[n\to\infty]{} \Phi(x)$ 几乎必然成立。因此，根据引理 2，Φ_n 几乎必然一致收敛于 Φ 在 K_ε 上。因此，以概率 1，对所有充分大的 n，$\inf_{x\in K_\varepsilon}\hat{\Phi}_n(x) > \Phi^* + 2\eta/3$。此外，同样以概率 1，对所有充分大的 n，$\hat{\Phi}_n(x^*) < \Phi^* + \eta/3$。

我们已经证明，以概率 1，对于所有充分大的 n，以下关系同时成立：$\hat{\Phi}_n(x) > \Phi_n(x^*)$。下面我们将证明，这一点结合 Φ_n 的凸性，意味着 Φ_n 的任意极小值点与 M^* 的距离至多为 ε。由此即可得到所需的结果。

假设存在一个极小值点 $\tilde{x}_n$ 满足 $d(\tilde{x}_n, M^*) \geq \varepsilon$。考虑一条测地线 $\gamma \in \Gamma_{x^*,\,\tilde{x}_n}$。那么，由于函数 $d(\cdot, M^*)$ 的连续性，必定存在某个 $t \in [0,1]$，使得 $\gamma(t) \in K_\varepsilon$。$\Phi_n$ 的凸性推出了 $\Phi_n \circ \gamma$ 的凸性，其最小值在 $t = 1$ 处取得。因此，$\Phi_n \circ \gamma$ 必定是非递增的，这意味着 $\inf_{x\in K_\varepsilon}\hat{\Phi}_n(x) \leq \hat{\Phi}_n(\gamma(t)) \leq \hat{\Phi}_n(x^*)$，从而得出矛盾。

□

3.2 黎曼框架

现在，我们证明，至少在黎曼流形中，为了保证 $\hat{x}_n$ 的一致性，可以去掉 ϕ 上的 Lipschitz 条件。本节假设 (M, g) 是完备黎曼流形，并且有以下定理。

THEOREM 2. 设 (M, g) 为黎曼流形。假设 M^* 非空且有界。对于 $n \geq 1$，令 $\hat{x}_n$ 为 Φ_n 的一个极小值点。则有 $d(\hat{x}_n, M^*) \xrightarrow[n\to\infty]{} 0$ 几乎必然。

同样，若 $\hat{x}_n$ 满足 $\Phi_n(\hat{x}_n) \leq \min_{x\in M}\Phi_n(x) + \varepsilon_n$，其中 ε_n 是任意一个几乎必然趋于零的非负误差项，当 $n \to \infty$ 时，该定理也适用。

一旦得到下面的引理 4 （该引理以引理 3为基础），定理 2 的证明就与定理 1的证明非常相似，因此我们省略证明。

LEMMA 3. *[27]* 黎曼流形上的任意凸函数都是连续的，且局部利普希茨。

通过改写 [47, 定理 10.8], 的证明，可以得到以下结果。

LEMMA 4. 设 $(f_n)_{n\geq 1}$ 是定义在 M 上的凸函数序列。假设对稠密子集 M_0 中的所有 x，都有 $f_n(x) \xrightarrow[n\to\infty]{} f(x)$，其中 $f : M \to \mathbb{R}$ 是给定的某个凸函数。那么，f_n 是一致利普希茨的，并且在 M 的任意紧子集上一致收敛到 f。

REMARK 1. 引理 *3* 在不假设 $\phi(z, \cdot)$ 对所有 L-Lipschitz 的情形下，是一致性证明的关键；对于所有 $z \in \mathcal{Z}$，存在某个 $L > 0$，且该值不依赖于 z。正如我们在上一节所指出的，我们不知道这个引理在更一般的完备测地空间中是否仍然成立，因此将此留作一个开放问题。

4. 渐近正态性

现在，我们转向 $\hat{x}_n$ 的渐近正态性。黎曼流形提供了一个自然且合理的框架，因为除了具有度量结构外，它们还具有光滑性，从而可以计算

一阶和二阶展开，并且可以在其切空间上定义高斯分布。因此，在本节中，我们假设 (M, g) 是一个完备黎曼流形。

THEOREM 3. 假设分布 P 和函数 ϕ满足以下假设：

- Φ 有唯一的极小值点 $x^* \in M$
- Φ在 x^* 处二阶可微，且 $\mathrm{H}\Phi(x^*)$ 是正定的。
- 存在 $\eta > 0$ 使得 $\mathbb{E}[\|g(Z,x)\|_{x^*}^2] < \infty$，对于所有满足 $d(x^*, x) \le \eta$的 $x \in M$，其中 $g(Z,x)$是 $\phi(Z,\cdot)$在点 x处的可测次梯度。

因此，$\hat{x}_n$ 是渐近正态的，即，

$$\sqrt{n} Log_{x^*}(\hat{x}_n) \xrightarrow[n\to\infty]{(d)} \mathcal{N}_{T_{x^*}M}(0, V(x^*))$$

其中 $V(x^*) = S(x^*)^{-1} B(x^*) S(x^*)^{-1}$，$S(x^*) = \mathrm{H}\Phi(x^*)$，$B(x^*) = \mathbb{E}[g(Z,x^*) g(Z,x^*)^\top]$ 并且我们用 $\mathrm{N}_{T_{x^*}M}(0, V(x^*))$ 表示欧几里得空间 $T_{x^*}M$上的正态分布，其均值为 0，协方差算子为 $V(x^*)$。

这里，$\mathrm{H}\Phi(x^*)$ 是 Φ 在点 x^* (参见附录B)。回忆一下，$\mathrm{Log}_{x^*}(\hat{x}_n)$ 是“$\hat{x}_n - x^*$”的黎曼类比，因此定理 3 是对$\hat{x}_n$ 在黎曼背景下渐近正态性的自然表述。

仔细阅读证明的结尾便会发现，在不改变该定理结论的前提下，可以假设$\hat{x}_n$ 不必是 Φ_n的极小值点，但应满足 $\hat{\Phi}_n(\hat{x}_n) \le \min_{x\in M} \hat{\Phi}_n(x) + o_P(n^{-1/2})$，其中 $o_P(n^{-1/2})$ 是一个可能为随机的误差项，它依概率以快于 $n^{-1/2}$的速率趋于零。

REMARK 2. 定理 3 的表述隐含地假设 ϕ 存在可测次梯度，即存在映射 $g : \mathsf{Z} \times M \to TM$，使得对于 P-几乎所有 $z \in \mathsf{Z}$ 以及所有 $x \in M$，$g(z,x) \in T_x M$ 是 $\phi(z,\cdot)$ 在点 $x \in M$ 处的次梯度；并且对于所有 $x \in M$，$g(\cdot, x) : \mathsf{Z} \to T_x M$ 都是可测的。根据关于可测选择的引理 8，这一点实际上是有保证的，见附录 A。还请注意，如果对于某个 $L \in L^2(P)$，$\phi(z,\cdot)$ 在 x^* 处局部满足 $L(z)$-Lipschitz 条件，且对 P-几乎所有 $z \in \mathsf{Z}$ 都如此，那么最后一个假设便自动成立。

定理 3 的证明深受 [28] 和 [44] 的启发，但需要运用精巧的工具来处理该问题的非欧几里得几何。证明的思路是说明

$$\hat{\Phi}_n(\mathrm{Exp}_{x^*}(u/\sqrt{n})) \approx \hat{\Phi}_n(x^*) + \frac{1}{\sqrt{n}} \langle u, n^{-1} \sum_{i=1}^n g(Z_i, x^*) \rangle_{x^*} + n^{-1} \langle u, \mathrm{H}\Phi(x^*) u \rangle_{x^*}$$

随着 n一致地增大时，在 $u \in T_{x^*}M$上成立。因此，左侧的极小值点根据定义即为 $\mathrm{Log}_{x^*}\hat{x}_n$，必然接近右侧的极小值点。右侧具有闭式解，并且由于中心极限定理而具有渐近正态性。主要困难在于处理 $\hat{\Phi}_n(\mathrm{Exp}_{x^*}(u/\sqrt{n}))$，它不是关于 $u \in T_{x^*}M$的凸函数，尽管 $\hat{\Phi}_n$在 M上是凸的。下面的引理 5 是我们最具技术性的结果，也是将 [28, 44] 中的技术推广到黎曼几何框架的关键论证。

LEMMA 5. 设 (M, g) 为一个黎曼流形，且 $x_0 \in M$。存在 $r > 0$，使得以下结论成立。设 $L > 0$。存在 $\alpha > 0$，使得对于所有在 $B(x_0, r)$ 上为 L-Lipschitz 的凸函数 $f : M \to \mathbb{R}$，映射

$$f \circ Exp_{x_0} + \frac{\alpha}{2} \|\cdot\|_{x_0}^2 : T_{x_0}M \to \mathbb{R}$$

在 $\{u \in T_{x_0}M : \|u\|_{x_0} \le r/2\}$ 上是凸的，其中 $\|\cdot\|_{x_0}$ 是由 g 在切空间 $T_{x_0}M$ 上诱导的范数。

在此引理的证明中，我们首先考虑 f 光滑的情形，然后通过用光滑函数逼近凸函数来处理一般情形，并利用[27, 定理 2]。

Proof of Theorem 3。固定 $u \in T_{x^*}M$。对于所有 $i = 1, \dots, n$，令

$$W_{i,n} = \phi(Z_i, \mathrm{Exp}_{x^*}(u/\sqrt{n})) - \phi(Z_i, x^*) - \frac{1}{\sqrt{n}}\langle u, g(Z_i, x^*)\rangle_{x^*}.$$

次梯度的定义表明 $W_{i,\ n} \ge 0$。令 π_n 为沿测地线 $\gamma(t) = \mathrm{Exp}_{x^*}(tu/\sqrt{n})$ $t \in [0,1]$从 x^* 到 $x_n := \mathrm{Exp}_{x^*}(u/\sqrt{n})$的平行移动。于是，$x^* = \mathrm{Exp}_{x_n}(-\pi_n^{-1}(u)/\sqrt{n})$。因此，可得：

$$\begin{aligned} W_{i,n} &\le \langle \pi_n(u)/\sqrt{n}, g(Z_i, x_n)\rangle_{x_n} - \langle u/\sqrt{n}, g(Z_i, x^*)\rangle_{x^*} \\ &= \langle u/\sqrt{n}, \pi_n^{-1}\left(g(Z_i, \mathrm{Exp}_{x^*}(u/\sqrt{n})\right) - g(Z_i, x^*)\rangle_{x^*}, \end{aligned} \tag{1}$$

其中，最后一个等号用到了 π_n 是等距映射这一事实。

因此，$\mathbb{E}[W_{i,n}^2] \le n^{-1}\mathbb{E}[Y_n^2]$，其中

$$Y_n = \langle u, \pi_n^{-1}\left(g(Z_1, \mathrm{Exp}_{x^*}(u/\sqrt{n}))\right) - g(Z_1, x^*)\rangle_x.$$

注意到，由于定理的最后一个假设，当 n 足够大时，$\mathbb{E}[Y_n^2]$ 是有限的。

上述不等式表明 $Y_n \ge 0$ 对所有 $n \ge 1$ 都成立；由引理 9可知，序列 $(Y_n)_{n\ge1}$ 单调不增。因此，Y_n 几乎必然收敛到一个非负随机变量 Y。现在我们来证明 $Y = 0$ 几乎必然成立。为此，我们来证明 $\mathbb{E}[Y_n] \xrightarrow[n\to\infty]{} 0$。由于根据单调收敛定理，$\mathbb{E}[Y_n]$ 必定收敛到 $\mathbb{E}[Y]$，因此我们将得到 $\mathbb{E}[Y] = 0$。结合 $Y \ge 0$ 几乎必然成立这一事实，即可得到所需结果。

固定 $x \in M$，使得 $d(x, x^*) \le \eta$。那么，对所有 $v \in T_xM$，$\phi(Z_1, \mathrm{Exp}_x(v)) \ge \phi(Z_1, x) + v, g(Z_1, x)x$。取期望可直接得到 $\mathbb{E}[g(Z_1, x)] \in \partial\Phi(x)$。因此，对于足够大的所有整数 n，使得 $\|u/\sqrt{n}\|_{x^*} \le \eta$，有 $g_n := \mathbb{E}[g(Z_1, \mathrm{Exp}_{x^*}(u/\sqrt{n})] \in \partial\Phi(\mathrm{Exp}_{x^*}(u/\sqrt{n}))$。因此，

$$\mathbb{E}[Y_n] = \langle u, \pi_n^{-1}(g_n) - \nabla\Phi(x^*)\rangle_{x^*} \xrightarrow[n\to\infty]{} 0$$

由于 Φ 在 x^* 处二阶可微。

最后，我们得到 $Y = 0$ 几乎必然。现在，$(Y_n)_{n\ge0}$ 是一个单调不增的非负序列，因此由单调收敛定理可知，$\mathbb{E}[Y_n^2]$ 在 $n \to \infty$时趋于零，从而使得

$$\mathrm{var}\left(\sum_{i=1}^n W_{i,n}\right) = \sum_{i=1}^n \mathrm{var}(W_{i,n}) \le \sum_{i=1}^n \mathbb{E}[W_{i,n}^2] \le \mathbb{E}[Y_n^2] \xrightarrow[n\to\infty]{} 0.$$

因此，由切比雪夫不等式，$\sum_{i=1}^n (W_{i,\ n} - \mathbb{E}[W_{i,n}])$ 依概率收敛到零，当$n \to \infty$时，也就是说，

$$\begin{aligned} &n\left(\hat{\Phi}_n(\mathrm{Exp}_{x^*}(u/\sqrt{n})) - \hat{\Phi}_n(x^*)\right) - \frac{1}{\sqrt{n}}\langle u, \sum_{i=1}^n g(Z_i, x^*)\rangle_{x^*} \\ &\quad - n\left(\Phi(\mathrm{Exp}_{x^*}(u/\sqrt{n})) - \Phi(x^*) - \langle u, \nabla\Phi(x^*)\rangle_{x^*}\right) \xrightarrow[n\to\infty]{} 0 \end{aligned}$$

依概率。由于 Φ 在 x^* 处二阶可微，我们得到

$$n\left(\hat{\Phi}_n(\mathrm{Exp}_{x^*}(u/\sqrt{n})) - \hat{\Phi}_n(x^*)\right) - \frac{1}{\sqrt{n}}\langle u, \sum_{i=1}^n g(Z_i, x^*)\rangle_{x^*} - \langle u, \mathrm{H}\Phi(x^*)u\rangle_{x^*} \xrightarrow[n\to\infty]{} 0 \tag{2}$$

依概率。对于 $n \geq 1$ 和 $u \in T_{x^*}M$，记为

$$F_n(u) = n\left(\hat{\Phi}_n(\mathrm{Exp}_{x^*}(u/\sqrt{n})) - \hat{\Phi}_n(x^*)\right) - \frac{1}{\sqrt{n}}\langle u, \sum_{i=1}^n g(Z_i, x^*)\rangle_{x^*}$$

并记为

$$F(u) = \langle u, \mathrm{H}\Phi(x^*)u\rangle_{x^*}.$$

我们已经证明，$F_n(u)$ 依概率收敛到 $F(u)$ 当 $n \to \infty$ 时，对任意固定的 $u \in T_{x^*}M$ 均成立。现在，我们希望利用引理 2。然而，无法保证函数 F_n 是凸函数。我们使用引理 5对这些函数进行调整。

令 $R > 0$ 为任意固定的正数。令 $K = B_{x^*}(0, R)$ 为 $T_{x^*}M$ 中以~原点为中心、半径为 R 的球，并令 $K = B(x^*, R)$。于是，K 和 K 均为紧集。根据大数定律，$\Phi_n(x)$ 几乎必然收敛到 $\Phi(x)$，对所有 $x \in K$ 均成立。因此，由引理 2 可知，$\sup_{x\in\tilde{K}} |\Phi_n(x) - \Phi(x)| \xrightarrow[n\to\infty]{} 0$ 几乎必然成立。于是，引理 4 保证存在一个（可能是随机的）$L > 0$，使得以下结论以概率 1 成立：

$$|\hat{\Phi}_n(x) - \hat{\Phi}_n(y)| \leq Ld(x, y), \forall n \geq 1, \forall x, y \in \tilde{K}.$$

现在，令 $r > 0$ 如引理 5所述。于是，存在一个（可能是随机的）$\alpha > 0$ 使得以概率 1，对所有 $n \geq 1$，$\hat{\Phi}_n \circ \mathrm{Exp}_{x^*} + \frac{\alpha}{2}\|\cdot\|^2_{x^*}$ 在 K 上是凸的。

对所有 $u \in K$ 和 $n \geq 1$，记 $\tilde{F}_n(u) = F_n(u) + \frac{\alpha}{2}\|u\|^2_{x^*}$ 和 $\tilde{F}(u) = F(u) + \frac{\alpha}{2}\|u\|^2_{x^*}$。对于所有充分大的 $n \geq 1$（使得 $R/\sqrt{n} \leq r$），每个 F_n 以概率 1 在 K 上是凸的，并且我们~显然有，对于所有 $u \in K$，$F_n(u) \xrightarrow[n\to\infty]{} F(u)$ 依概率成立。因此，由引理 2 可得~$\sup_{u\in K} |F_n(u) - F(u)| = \sup_{u\in K} |F_n(u) - F(u)| \xrightarrow[n\to\infty]{} 0$ 依概率成立。

最后，我们已经证明，对于任意 $R > 0$，

$$\sup_{\|u\|_{x^*}\leq R} \left| n\left(\hat{\Phi}_n(\mathrm{Exp}_{x^*}(u/\sqrt{n})) - \hat{\Phi}_n(x^*)\right) - \frac{1}{\sqrt{n}}\langle u, \sum_{i=1}^n g(Z_i, x^*)\rangle_{x^*} - \langle u, \mathrm{H}\Phi(x^*)u\rangle_{x^*} \right| \xrightarrow[n\to\infty]{} 0 \tag{3}$$

依概率。

根据定义，$\hat{x}_n$, $\Phi_n(\mathrm{Exp}_{x^*}(u/\sqrt{n}))$ 在 $u = \sqrt{n}\mathrm{Log}_{x^*}(\hat{x}_n)$ 处取最小值。此外，很容易验证，$\frac{1}{\sqrt{n}}$ $u, \sum_{i=1}^n g(Z_i, x^*)_{x^*} - u, \mathrm{H}\Phi(x^*)u_{x^*}$ 在以下位置取最小值。

$$u = -\mathrm{H}\Phi(x^*)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n g(Z_i, x^*).$$

注意，$g(Z_1, x^*) \ldots, g(Z_n, x^*)$ 是 $T_{x^*}M$ 中独立同分布（i.i.d.）的随机向量。根据一阶条件，由于$\mathbb{E}[g(Z_1, x^*)] = \nabla\Phi(x^*) = 0$，它们的均值为零；同时，根据定理的最后一个假设，它们具有二阶矩。因此，多元中心极限定理表明，

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n g(Z_i, x^*) \xrightarrow[n\to\infty]{} \mathcal{N}_{T_{x^*}M}(0, B(x^*)) \tag{4}$$

依分布。特别地，序列 $\frac{1}{\sqrt{n}}\sum_{i=1}^n g(Z_i, x^*)$ 依概率有界，因此可以在 $(R$（3）中选取 $\frac{1}{\sqrt{n}}\sum_{i=1}^n g(Z_i, x^*) \in B_{x_0}(0, R)$，使得对所有 $n \geq 1$，该结论以任意大的概率成立。

由 (3) 易得，$\sqrt{n}\mathrm{Log}_{x^*}(\hat{x}_n) - \mathrm{H}\Phi(x^*)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n g(Z_i, x^*)$ 依概率收敛于零，当 $n \to \infty$ 时。因此，根据斯卢茨基定理，$\sqrt{n}\mathrm{Log}_{x^*}(\hat{x}_n)$ 的分布极限与 $\mathrm{H}\Phi(x^*)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n g(Z_i, x^*)$ 相同，结合 (4) 即可完成证明。

□

Proof of Lemma 5. 固定 $r > 0$ （待定），并记为 $K = B(x_0, r)$；根据霍普夫-里诺定理 [21, 第 7 章, 定理 2.8]，它是紧的。令 $f : M \to \mathbb{R}$ 在 K 上为凸且满足 L-Lipschitz 条件（$L > 0$），并设 $F = f \bigcirc \mathrm{Exp}_{x_0} : T_{x_0}M \to \mathbb{R}$。

我们将此引理的证明分为几个步骤。

步骤 1： f 为光滑的情形。首先，假设 f 是光滑的。那么，F 是一个光滑函数。¯下文中，我们用 $\mathrm{H}f$ 表示 f 的海森矩阵，用 $\mathrm{H}F$ 表示 F 的海森矩阵（也就是说，对于定义在 M 上的函数的海森矩阵，我们记作 H；对于定义在切空间$T_{x_0}M$上的函数的海森矩阵，我们也记作 H）。固定 $u, v \in T_{x_0}M$，并令 $\phi(t) = F(u + tv)$，其中 $s \in \mathbb{R}$。记 $\eta(s,t) = \mathrm{Exp}_{x_0}(t(u + sv))$，对所有 $s, t \in \mathbb{R}$均成立。 F的海森矩阵为 $\bar{\mathrm{H}}F(u)(v,v) = \phi''(0)$。对于所有 $s \in \mathbb{R}$,

$$\phi'(s) = \langle \nabla f(\eta(s,1)), \frac{\partial \eta}{\partial s}(s,1)\rangle_{\eta(s,1)}$$

and

$$\phi''(s) = \langle \mathrm{H}f(\eta(s,t))\frac{\partial \eta}{\partial s}(s,1), \frac{\partial \eta}{\partial s}(s,1)\rangle_{\eta(s,1)} + \langle \nabla f(\eta(s,1)), D_s\frac{\partial \eta}{\partial s}(s,1)\rangle_{\eta(s,1)} \tag{5}$$

其中 D_s 是沿路径的协变导数 $\eta(\cdot,1)$。注意，(5)右侧的第一项非负，这是由于 f 的凸性（参见 [52, 定理 6。2]）。因此，

$$\begin{aligned}\phi''(0) &\geq \langle \nabla f(\eta(0,1)), D_s\frac{\partial \eta}{\partial s}(0,1)\rangle_{\eta(0,1)} \geq -\|\nabla f(\eta(0,1))\|\|D_s\frac{\partial \eta}{\partial s}(0,1)\| \\ &\geq -L\|D_s\frac{\partial \eta}{\partial s}(0,1)\| \end{aligned} \tag{6}$$

其中第二个不等式由柯西-施瓦茨不等式得出，最后一个不等式则由假设 f 在 K 上是 L-Lipschitz 得出。

对于所有 w、$w' \in T_{x_0}M$，令 $\Gamma_{w,\, w'}(s,t) = \mathrm{Exp}_{x_0}(t(w + sw'))$ $s, t \in [0,1]$。特别地，$\eta = \Gamma_{u,\, v}$。于是，$D_s \frac{\partial \Gamma_{u,\, v}}{\partial s}(0,1) = \|v\|^2 D_s \frac{\partial \Gamma_{u,\, \tilde{v}}}{\partial s}(0,1)$，其中 $\tilde{v} = \frac{v}{\|v\|_{x_0}}$。由于映射 $w, w' \mapsto D_s \frac{\partial \Gamma_{w,\, w'}}{\partial s}(0,1)$ 连续，存在一个常数 $C > 0$ （它仅取决于 r 以及 (M,g) 的几何），使得

$$\|D_s\frac{\partial \Gamma_{w,w'}}{\partial s}(0,1)\| \leq C,$$

对所有 u，$v \in T_{x_0}M$其中 $\|u\|_{x_0} \leq r$ 且 $\|v\|_{x_0} \leq 1$。因此，(6) 得出 $\phi''(0) \geq -LC\|v\|_{x_0}^2$。因此，令 $\alpha_1 = LC$，我们证明了$F + \frac{\alpha_1}{2}\|\cdot\|_{x_0}^2$在$\{u \in T_{x_0}M : \|u\|_{x_0} \leq r\}$上是凸的。

步骤 2：不再假设 f 是光滑函数的情形。

步骤 2.1： f 的光滑逼近。对所有正整数 n，令 $f_n : M \to \mathbb{R}$ 为由下式定义的函数

$$f_n(x) = n^d \int_{T_xM} f(\mathrm{Exp}_x(v))k(n\|v\|_x)\,\mathrm{d}\Omega_x(v)$$

其中， $k : \mathbb{R} \to \mathbb{R}$是一个支撑于 $[0,1]$上的非负光滑函数，满足 $\int_{\mathbb{R}} k(t)\mathrm{d}t = 1$，且 Ω_x是黎曼体积经指数映射 Exp_x推前到 T_xM 上的推前测度。

根据[27, 定理 2], 每个 f_n 都是光滑的，并且序列$(f_n)_{n\geq 1}$ 收敛到 f 在 K上一致收敛，并且满足

$$\liminf_{n\to\infty} \inf_{x\in K} \lambda_{\min}(\mathrm{H}f_n(x)) \geq 0,$$

其中 $\lambda_{\min}$ 表示最小特征值。固定 $\varepsilon > 0$，之后我们会将其选得足够小。不失一般性，假设对所有 $n \geq 1$ 都有 $\inf_{x\in K} \lambda_{\min}(\mathrm{H}f_n(x)) \geq -\varepsilon$（因为不妨忽略序列 $(f_n)_{n\geq 1}$的前几项）。

步骤 2.2： f_n在 $B(x_0, r/2)$ 上是 Lipschitz 的。下面证明，只要 n 足够大，f_n在 $B(x_0, r/2)$ 上就是 Lipschitz 的。令 $x, y \in B(x_0, r/2)$，并令 $\pi_{x,y} : T_xM \to T_yM$为从 x 到 y的平行移动映射。由于$\pi_{x,y}$ 是等距映射，它将测度 Ω_x 映射到 Ω_y，因此有

$$f_n(y) = n^d \int_{T_xM} f(\mathrm{Exp}_y(\pi_{x,y}(v)))k(n\|v\|_x)\,\mathrm{d}\Omega_x(v).$$

因此,

$$\begin{aligned} |f_n(x) - f_n(y)| &= n^d \left| \int_{T_xM} (f(\mathrm{Exp}_x(v)) - f(\mathrm{Exp}_y(\pi_{x,y}(v))))k(n\|v\|_x) \right| \mathrm{d}\Omega_x(v) \\ &\leq n^d \int_{T_xM} \left| f(\mathrm{Exp}_x(v)) - f(\mathrm{Exp}_y(\pi_{x,y}(v))) \right| k(n\|v\|_x)\,\mathrm{d}\Omega_x(v) \\ &\leq Ln^d \int_{T_xM} d(\mathrm{Exp}_x(v), \mathrm{Exp}_y(\pi_{x,y}(v))k(n\|v\|_x)\,\mathrm{d}\Omega_x(v) \end{aligned} \tag{7}$$

只要 n 足够大（即 $n \geq 2/r$），最后一个不等式便成立，因为 f 在$K = B(x_0, r)$ 上是 L-Lipschitz 的，并且在最后一个积分中，只有满足 $\|v\|_{x_0} \leq n^{-1} \leq r/2$ 的向量 $v \in T_xM$ 对积分有贡献。最后，映射 $(x, y, v) \mapsto \mathrm{Exp}_y(\pi_{x,y}(v))$ 是光滑的，因此在任意紧集上都是 Lipschitz 的，所以存在一个与 x 和 y 无关的 $C > 0$，使得对所有满足$\|v\|_{x_0} \leq r/2$ 的 $v \in T_xM$，都有 $d(\mathrm{Exp}_x(v), \mathrm{Exp}_y(\pi_{x,y}(v))) \leq Cd(x, y)$。因此，由 (7) 得

$$|f_n(x) - f_n(y)| \leq LCd(x, y)n^d \int_{T_xM} k(n\|v\|_x)\,\mathrm{d}\Omega_x(v) = LCd(x, y), \tag{8}$$

对于所有 $x, y \in B(x_0, r/2)$。因此，对于所有整数 $n \geq 2/r$， f_n 在 $B(x_0, r/2)$ 上是 CL-Lipschitz 的。

步骤 2.3：向每个 f_n 添加一个固定函数，使它们成为凸函数。 假设 r 和 ε 按照下述引理 ϕ 6 进行选取，并令 ϕ 为该引理中给出的函数。则，对于所有 $n \geq 1$，$f_n + \phi$ 在 K上是凸函数，因为其海森矩阵在任意 $x \in K$处都是半正定的。

步骤 2.4：将光滑情形（步骤 1）应用于每个 $f_n + \phi$ 在 $B(x_0, r/2)$ 上。对于所有 $n \geq 1$， $f_n + \phi$ 都是光滑的。此外， f_n 在 $B(x_0, r/2)$ 上是 CL-Lipschitz 的，其中 $C > 0$ 不依赖于 f；而 ϕ 是光滑的，因此它在 $B(x_0, r/2)$ 上是 L'-Lipschitz 的，其中存在某个 $L' > 0$，且它也不依赖于 f。因此，应用步骤 1 中证明的结果，我们得到存在一个不依赖于 f 的 $\alpha_2 > 0$，使得对于所有足够大的整数 n， $f_n \bigcirc \mathrm{Exp}_{x_0} + \phi \bigcirc \mathrm{Exp}_{x_0} + \frac{\alpha_2}{2}\|\cdot\|^2_{x_0}$ 在 $\{u \in T_{x_0}M : \|u\|_{x_0} \leq r/2\}$ 上是凸函数。令 $n \to \infty$ 取极限，即可得到 $f \bigcirc \mathrm{Exp}_{x_0} + \phi \bigcirc \mathrm{Exp}_{x_0} + \frac{\alpha_2}{2}\|\cdot\|^2_{x_0}$ 在 $\{u \in T_{x_0}M : \|u\|_{x_0} \leq r/2\}$ 上是凸函数。

由于 ϕ 是光滑的，因此 $\phi \bigcirc \mathrm{Exp}_{x_0}$ 也是光滑的，所以存在 $\beta > 0$，使得 $\lambda_{\max}(\bar{\mathrm{H}}(\phi \bigcirc \mathrm{Exp}_{x_0})(u)) \leq \beta$，对所有满足 $\|u\|_{x_0} \leq r/2$ 的 $u \in T_{x_0}M$ 均成立。因此，通过令 $\alpha_3 = \alpha_2 + \beta$，我们得到 $f \bigcirc \mathrm{Exp}_{x_0} + \frac{\alpha_3}{2}\|\cdot\|^2_{x_0}$ 在 $\{u \in T_{x_0}M : \|u\|_{x_0} \leq r/2\}$ 上是凸的。

步骤 3：结论。现在，我们可以通过取 $\alpha = \max(\alpha_1, \alpha_3)$ 将所有情形合并。

□

LEMMA 6. 设 (M,g) 是一个无边界的完备黎曼流形，且 $x_0 \in M$。存在 $r,\varepsilon > 0$ 以及一个光滑函数 $\phi : M \to \mathbb{R}$，使得 $\lambda_{\min}(\mathrm{H}\phi(x)) \geq \varepsilon$ 对所有 $x \in B(x_0,r)$ 均成立。

PROOF. 设 r_1 满足 $B(x_0,r_1)$ 是单连通的。对于所有 $x \in M$，以及所有满足 $\|u\|_x = \|v\|_x = 1$ 的线性无关向量 $u,v \in T_xM$，令 $\kappa(x,u,v)$ 表示 M 在 x 处沿由 u 和 v 张成的二维平面的截面曲率（参见 [21, 命题 3.1 和定义 3.2]）。于是，κ 是光滑的，因此存在 $\kappa_0 < \infty$，它是 M 在紧集 $x \in B(x_0,r_1)$ 上所有点处截面曲率的上界。因此，由引理 12 可知，$B(x_0,r_1)$ 是 CAT(κ_0) 空间。取 $r = \min(r_1, D_{\kappa_0}/4)$ 以及 $\phi = \frac{\varepsilon}{2\alpha_{2r},\kappa_0} d(\cdot,x_0)^2$，，其中 D_{κ_0} 和 α_{2r,κ_0} 在引理 11 中定义，参见附录 D.2，即可得到所需结果。

□

参考文献

[1] Eddie Aamari 和 Cl´ement Levrard。流形、切空间与曲率估计的非渐近速率。2019。[2] Martial Agueh 和 Guillaume Carlier。瓦瑟斯坦空间中的重心。《SIAM 数学分析杂志》，43(2):904–924，2011。[3] Adil Ahidar-Coutrix、Thibaut Le Gouic 和 Quentin Paris。度量空间中经验重心的收敛速率：曲率、凸性与可延拓测地线。《概率论及相关领域》，177(1-2):323–368，2020。[4] Stephanie Alexander、Vitali Kapovitch 和 Anton Petrunin。Alexandrov 几何：初步版本第 1 号。arXiv 电子预印本，页码 arXiv–1903，2019。[5] Jason M. Altschuler 和 Enric Boix-Adsera。固定维度下可在多项式时间内计算 Wasserstein 重心。《机器学习研究杂志》，22:44–1，2021。[6] Jason M. Altschuler 和 Enric Boix-Adsera。Wasserstein 重心的计算是 NP-hard 的。《SIAM 数据科学数学杂志》，4(1):179–203，2022。[7] Kimon Antonakopoulos、Elena Veronica Belmega 和 Panayotis Mertikopoulos。超越 Lipschitz 连续性的在线与随机优化：一种黎曼方法。载于 ICLR 2020——国际学习表征会议，页码 1–20，2020。[8] Rajendra Bhatia。正定矩阵。载于《正定矩阵》。普林斯顿大学出版社，2009。[9] Rajendra Bhatia 和 John Holbrook。黎曼几何与矩阵几何均值。《线性代数及其应用》，413(2-3):594–618，2006。[10] Rajendra Bhatia、Tanvi Jain 和 Yongdo Lim。关于正定矩阵之间的 Bures-Wasserstein 距离。《数学论述》，37(2):165–191，2019。[11] Rabi Bhattacharya 和 Lizhen Lin。关于 Fr´echet 均值的综合性中心极限定理与非参数推断：非欧几里得空间。《美国数学会会刊》，145(1):413–428，2017。[12] Rabi Bhattacharya 和 Vic Patrangenaru。流形上内禀与外在样本均值的大样本理论。《统计年刊》，31(1):1–29，2003。[13] Rabi Bhattacharya 和 Vic Patrangenaru。流形上内禀与外在样本均值的大样本理论：II。《统计年刊》，页码 1225–1259，2005。[14] Louis J. Billera、Susan P. Holmes 和 Karen Vogtmann。系统发育树空间的几何。《应用数学进展》，27(4):733–767，2001。[15] Martin R Bridson 和 Andr´é Haefliger。非正曲率度量空间，第 319 卷。Springer Science & Business Media，2013。[16] Dmitri Burago、Yuri Burago 和 Sergei Ivanov。度量几何教程，第 33 卷。美国数学会，2001。

[17] Yuan Shih Chow 和 Henry Teicher。概率论：独立性、可交换性与鞅。Springer Science & Business Media，2003。[18] Sebastian Claici、Edward Chien 和 Justin Solomon。随机 Wasserstein 重心。载于国际机器学习会议，页码 999–1008。PMLR，2018。[19] Christopher Criscitiello 和 Nicolas Boumal。流形上非凸优化的加速一阶方法。《计算数学基础》，页码 1–77，2022。[20] Marco Cuturi 和 Arnaud Doucet。Wasserstein 重心的快速计算。载于国际机器学习会议，页码 685–693。PMLR，2014。[21] Manfredo Perdigao Do Carmo。黎曼几何，第 6 卷。Springer，1992。[22] Benjamin Eltzner、Fernando Galaz-Garcia、Septhan F. Huckemann 和 Wilderich Tuschmann。割迹的稳定性与黎曼流形 Fr´echet 均值的中心极限定理。arXiv 预印本 arXiv:1909.00410，2019。[23] Benjamin Eltzner 和 Stephan F Huckemann。流形的涂抹型中心极限定理及其在高维球面中的应用。《统计年刊》，47(6):3360–3381，2019。[24] Maurice Fr´echet。度量空间中任意性质的随机元素。Ann. I nst. H. Poincar´é，10:215–310，1948。[25] Kei Funano。取值于 $\mathbb{R}$-树与 Hadamard流形的随机过程的收敛速率。《大阪数学杂志》，47(4):911–920，2010。[26] Christopher R Genovese、Marco Perone Pacifico、Verdinelli Isabella、Larry Wasserman 等。流形极小极大估计。《机器学习研究杂志》，13:1263–1291，2012。[27] Robert E. Greene 和 Hung Wu。测地凸函数的次调和性与多重次调和性。《印第安纳大学数学杂志》，22(7):641–653，1973。[28] Shelby J. Haberman。凹性与估计。《统计年刊》，页码 1631–1661，1989。[29] CH Himmelberg、T Parthasarathy 和 FS Van Vleck。关于可测关系。《数学基础》，61:161–167，1982。[30] Thomas Hotz、Sean Skwerer、Stephan Huckemann、Huiling Le、J Stephen Marron、Jonathan C Mattingly、Ezra Miller、James Nolen、Megan Owen 和 Vic Patrangenaru。开放书页空间上的黏滞型中心极限定理。《应用概率年刊》，2013。[31] Peter J Huber。位置参数的稳健估计。载于《统计学的突破：方法论与分布》，页码 492–518，1992。[32] Stephan F. Huckemann。平面形状平均测地线的内禀推断与通过叶片生长进行树判别。《统计年刊》，39(2):1098–1124，2011。[33] Hermann Karcher。黎曼质心与磨光函数平滑。《纯粹与应用数学通讯》，30(5):509–541，1977。[34] David George Kendall、Dennis Barden、Thomas K. Carne 和 Huiling Le。形状与形状理论，第 500 卷。John Wiley & Sons，2009。[35] Wilfrid S. Kendall 和 Huiling Le。独立且非同分布的流形值随机变量的经验 Fr´echet 均值的极限定理。《巴西概率与统计杂志》，25(3):323 – 352，2011。[36] Alexey Kroshnin、Vladimir Spokoiny 和 Alexandra Suvorikova。Bures-Wasserstein 重心的统计推断。《应用概率年刊》，31(3):1264–1298，2021。[37] Alexey Kroshnin、Nazarii Tupitsa、Darina Dvinskikh、Pavel Dvurechensky、Alexander Gasnikov 和 Cesar Uribe。关于 Wasserstein 重心近似的复杂度。载于国际机器学习会议，页码 3530–3540。PMLR，2019。[38] Thibaut Le Gouic 和 Jean-Michel Loubes。Wasserstein 重心的存在性与一致性。《概率论及相关领域》，168(3):901–917，2017。[39] Thibaut Le Gouic、Quentin Paris、Philippe Rigollet 和 Austin J Stromme。快速收敛

关于亚历山德罗夫空间和瓦瑟斯坦空间中的经验重心。《欧洲数学会杂志》，2022。[40] John M. Lee。《光滑流形导论》。Springer，2012。[41] John M. Lee。《黎曼流形导论》，第2卷。Springer，2018。[42] Yongdo Lim 和 Mikl´os P´alfia。哈达玛空间上用于最小二乘均值的加权确定性游走。《伦敦数学会通报》，46，2014年5月。[43] Estelle Massart、Julien M. Hendrickx 和 P-A Absil。赋予 Bures-Wasserstein 度量的固定秩半正定矩阵流形的曲率。载于《信息几何科学：第四届国际会议 GSI 2019，法国图卢兹，2019年8月27–29日，会议论文集》，第739–748页。Springer，2019。[44] Wojciech Niemiro。通过凸极小化定义的 M估计量的渐近性质。《统计年刊》，第1514–1533页，1992。[45] Shin-ichi Ohta。度量空间的凸性。《Geom. Dedicata》，125：225–250，2007。[46] Shin-ichi Ohta 和 Mikl´os P´alfia。亚历山德罗夫空间中的离散时间梯度流与大数定律。《Calc. Var. Partial Differential Equations》，54(2)：1591–1610，2015。[47] R Tyrrell Rockafellar。《凸分析》，第18卷。普林斯顿大学出版社，1970。[48] Rolf Schneider。《凸体：布伦-闵可夫斯基理论》。第151号。剑桥大学出版社，2014。[49] Christof Sch¨otz。通过四点不等式得到广义 Fr´échet 均值的收敛速率。《电子统计学杂志》，13：4280–4345，2019。[50] Nicolas Schreuder、Victor-Emmanuel Brunel 和 Arnak Dalalyan。无支配条件下生成模型的统计保证。载于《算法学习理论》，第1051–1071页。PMLR，2021。[51] Karl-Theodor Sturm。非正曲率度量空间上的概率测度。载于《流形、图和度量空间上的热核与分析》（巴黎，2002），《当代数学》第338卷，第357–390页。美国数学会，普罗维登斯，罗得岛州，2003。[52] Constantin Udriste。《黎曼流形上的凸函数与优化方法》，第297卷。Springer Science & Business Media，2013。[53] Aad W Van Der Vaart 和 Jon A Wellner。《弱收敛》。Springer，1996。[54] Daniel H. Wagner。可测选择定理综述。《SIAM控制与优化杂志》，15(5)：859–903，1977。[55] Hongyi Zhang 和 Suvrit Sra。测地凸优化的一阶方法。载于《学习理论会议》，第1617–1638页。PMLR，2016。[56] Hongyi Zhang 和 Suvrit Sra。迈向黎曼加速梯度方法。arXiv 预印本 arXiv:1806.02812，2018。[57] Herbert Ziezold。关于拟度量空间中随机元素的期望图形与强大数定律。载于《第七届布拉格信息论、统计决策函数、随机过程会议暨1974年欧洲统计学家会议论文集》，第591–602页。Springer，1977。

附录 A：关于次梯度的可测选择

LEMMA 7. 设 $(\mathbb{Z}, \mathcal{F}, P)$ 是一个概率空间，(M, g) 是一个无边界的黎曼流形。设 $\phi : \mathbb{Z} \times M \to \mathbb{R}$ 是一个函数，使得

- $\phi(z, \cdot)$ 对 P-几乎所有 $z \in \mathbb{Z}$ 都是凸的；
- $\phi(\cdot, x)$ 对所有 $x \in M$ 都是可测的。

那么，存在一个映射 $g : \mathbb{Z} \times M \to TM$ 使得

- $g(z, x) \in T_x M$ 是 $\phi(z, \cdot)$ 在 x 处的次梯度，对所有 $x \in M$，且对 P-几乎所有 $z \in \mathbb{Z}$；
- $g(\cdot, x)$ 对所有 $x \in M$ 都是可测的。

PROOF. 设 $N \in \mathcal{F}$ 满足 $P(N) = 0$，并且 $\phi(z, \cdot)$ 对所有 $z \in \mathbb{Z} \setminus N$ 都是凸函数。令 $\tilde{\mathbb{Z}} = \mathbb{Z} \setminus N$。只需证明，对于所有 $x \in M$，存在一个定义在 M 上的可测函数 $g(\cdot, x)$，使得对于所有 $z \in Z$，$g(z, x)$ 是 $\phi(z, \cdot)$ 在 x 处的一个次梯度。

固定 $x_0 \in M$。对于所有 $z \in \mathbb{Z}$，令 $\Gamma(z)$ 为 $\phi(z, \cdot)$ 在点 x_0 处的所有次梯度组成的集合。根据 [52, 定理 4.5]，$\Gamma(z) \neq \varnothing$，对所有 $z \in \mathbb{Z}$ 均成立。因此，只需证明 Γ 在 $\mathbb{Z}$ 上的限制 $\tilde{\Gamma}$ 是弱可测的，即对于所有开集 $U \subseteq T_{x_0} M$，都有 $\{z \in \mathbb{Z} : \Gamma(z) \cap U \neq \varnothing\} \in \mathcal{F}$（定义见[29]）。

可以将 $\Gamma(z)$ 写成

$$\begin{aligned}\Gamma(z) &= \bigcap_{v \in T_{x_0} M} \{u \in T_{x_0} M : \phi(z, \operatorname{Exp}_{x_0}(v)) \geq \phi(z, x_0) + \langle u, v \rangle_{x_0}\} \\ &= \bigcap_{v \in E_0} \{u \in T_{x_0} M : \phi(z, \operatorname{Exp}_{x_0}(v)) \geq \phi(z, x_0) + \langle u, v \rangle_{x_0}\},\end{aligned}$$

其中，E_0 是 $T_{x_0} M$ 的某个固定的可数稠密子集。最后一个等式由 $\phi(z, \cdot)$ 的连续性得出，见引理 3。对于所有 $v \in E_0$，令 $\Gamma_v(z) = \{u \in T_{x_0} M : \phi(z, \operatorname{Exp}_{x_0}(v)) \geq \phi(z, x_0) + u, v_{x_0}\}$。

由于可数多个 $\Gamma_v(z)$ $v \in E_0$ 中的每一个都是闭集，[29, 推论 4.2] 表明，只需证明每个 Γ_v 都是弱可测的。固定 $v \in E_0$，并令 U 为 $T_{x_0} M$ 的一个开子集。于是，对于所有 $z \in \mathbb{Z}$，

$$\begin{aligned}\Gamma_v(z) \cap U \neq \varnothing &\iff \exists u \in U, \phi(z, \operatorname{Exp}_{x_0}(v)) \geq \phi(z, x_0) + \langle u, v \rangle_{x_0} \\ &\iff \exists u \in U \cap E_0, \phi(z, \operatorname{Exp}_{x_0}(v)) \geq \phi(z, x_0) + \langle u, v \rangle_{x_0}.\end{aligned}$$

最后一个等价关系源于以下事实：由于 U 是开集，$U \cap E_0$ 在 U 中稠密；并且如果某个 $u \in U$ 满足 $\phi(z, \operatorname{Exp}_{x_0}(v)) \geq \phi(z, x_0) + u, v_{x_0}$，那么，对于任意范数足够小、满足 $v, v'_{x_0} \leq 0$ 的 $v' \in T_{x_0} M$，任何形如 $u' = u - v'$ 的 $u' \in U$ 都满足条件。至少存在一个这样的 u' 落在 $U \cap E_0$ 中。

最终，得到：

$$\{z \in \tilde{\mathcal{Z}} : \Gamma_v(z) \cap U \neq \varnothing\} = \bigcup_{u \in E_0} \{z \in \tilde{\mathcal{Z}} : \phi(z, \operatorname{Exp}_{x_0}(v)) \geq \phi(z, x_0) + \langle u, v \rangle_{x_0}\},$$

其属于 $\mathcal{F}$，因为假设对所有 $x \in M$，$\phi(\cdot, x)$ 都是可测的。

<escape>利用引理 8即可得出结论。</escape>

□

LEMMA 8. 设 $(\mathbb{Z}, \mathcal{F}, P)$ 是一个概率空间，E是一个波兰空间。设 Γ 是一个函数，它将任意 $z \in \mathbb{Z}$ 映射到 E的一个闭子集。假设存在 $N \in \mathcal{F}$，使得 $P(N) = 0$

并且对所有 $z \in \mathsf{Z}$, $\Gamma(z) \neq \varnothing$。进一步假设，对于 E的所有开子集 U，$\{z \in \mathsf{Z} : N : \Gamma(z) \cap U \neq \varnothing\} \in \mathsf{F}$。那么，存在一个函数 $g : \mathsf{Z} \to E$，使得 $g(z) \in \Gamma(z)$，对P-几乎所有 $z \in \mathsf{Z}$ 都成立。

Proof. 考虑由以下方式定义的概率空间 $(\mathsf{Z}, \mathsf{F}, P)$:

- $\mathsf{Z} = \mathsf{Z} \setminus N$;
- $\mathsf{F} = \{F \setminus N : F \in \mathsf{F}\}$;
- P作为 P在 (Z, F) 上的限制。

那么，对于所有 $z \in \mathcal{Z}$, $\Gamma(z) \neq \varnothing$ 因此，借助 [54, 定理 4。1], 可以找到一个可测选择g，它是 Γ 在 $\mathcal{Z}$上的限制。令 $g(z)$ 在 E 上取任意常数值 $z \in N$，即可得到一个函数 $g : \mathcal{Z} \to E$，满足所需条件。

□

附录 B：黎曼流形与测地凸性的基本要素

在本节中，设 (M, g) 为一个完备的 d 维黎曼流形。也就是说，M 是一个 d 维光滑流形，而黎曼度量 g 为每个切空间 $T_xM, x \in M$ 配备了内积 $g_x(u, v)$, $u, v \in T_xM$。此外，根据霍普夫-里诺定理 [21, 第 7 章，定理 2。8], (M, d) 的完备性确保任意两点 $x, y \in M$ 之间至少存在一条最短测地线，即一条光滑路径 $\gamma : [0, 1] \to M$，使得 $\gamma(0) = x, \gamma(1) = y$，并且 $d(\gamma(s)\ \gamma(t)) = |s - t| d(x, y)$，对所有 $s, t \in [0, 1]$ 均成立。

B.1 关于光滑函数及其导数

如果 $f : M \to \mathbb{R}$是一个光滑函数，则其梯度是映射 ∇f，满足对所有 $x \in M$, $\nabla f(x) \in T_xM$，并且对所有 $u \in T_xM$,

$$\langle \nabla f(x), u \rangle_x = (f \circ \gamma)'(0)$$

其中，$\gamma : [0, 1] \to M$ 是任意满足 $\gamma'(0) = u$的光滑路径。f 在任意 $x \in M$ 处的海森矩阵由线性映射 $\mathrm{H}f(x) : T_xM \to T_xM$表示，满足对所有 $u \in T_xM$,

$$\langle u, \mathrm{H}f(x)u \rangle_x = (f \circ \gamma)''(0)$$

其中，γ 是满足 $\gamma(0) = x$、$\gamma'(0) = u$的测地线路径。

按照我们的记号，任意光滑函数 $f : M \to \mathbb{R}$ 在给定点 $x \in M$ 处都有如下形式的二阶泰勒展开:

$$f(\mathrm{Exp}_x(tu)) = f(x) + t\langle \nabla f(x), u \rangle_x + \frac{t^2}{2} \langle u, \mathrm{H}f(x)u \rangle_x + o(t^2)$$

当 $t \to 0$时，对于任意固定的 $u \in T_xM$。

B.2 关于凸函数

回忆一下，在 M上的凸函数是指满足 $f(\gamma(t)) \leq (1-t)f(x) + tf(y)$ 的任意函数 $f : M \to \mathbb{R}$，对于所有的 $t \in [0, 1]$、x、$y \in M$，以及从x 到 y的所有常速最短测地线 $\gamma : [0, 1] \to M$均成立。正如在欧几里得空间中一样，凸函数自动连续 [52, 定理 3.6]（注意，我们只考虑整个空间 M上的凸函数，并假设该空间没有边界）。次梯度的概念也可以扩展到黎曼情形。

Definition 2. 令 $f : M \to \mathbb{R}$ 为一个凸函数，且 $x \in M$。向量 $v \in T_xM$ 是 M 在 x 处的次梯度，当且仅当

$$f(Exp_x(u)) \geq f(x) + \langle v, u \rangle_x,$$

对所有 $u \in T_x(M)$。

在 f 处 $x \in M$ 的所有次梯度集合记作 $\partial f(x)$ 且 [52, 定理 4。5] 保证 $\partial f(x) \neq \varnothing$对所有 $x \in M$ (再次注意，我们只考虑 M 没有边界的情形)。为简便起见，这里与 [52], 不同，次梯度是切空间中的元素，而不是其对偶空间中的元素。

凸函数的次梯度满足以下单调性。

Lemma 9. 设 $f : M \to \mathbb{R}$ 为凸函数。设 $x \in M$, $u \in T_xM$ 并令 $y = \mathrm{Exp}_x(u)$。令$g_1 \in \partial f(x)$ 且 $g_2 \in \partial f(y)$。那么，

$$\langle u, \pi_{x,y}^{-1}(g_2) - g_1 \rangle_x \geq 0,$$

其中 $\pi_{x,\ y}$ 是沿由 $\gamma(t) = \mathrm{Exp}_x(tu)$ $t \in [0,1]$给出的测地线，从 x 到 y 的平行移动。

Proof. 首先注意，令 $v = \pi_{x,y}(u)$，则有 $x = \mathrm{Exp}_y(-v)$。根据次梯度的定义，可得以下不等式:

$$f(y) \geq f(x) + \langle u, g_1 \rangle_x$$

and

$$f(x) \geq f(y) + \langle -v, g_2 \rangle_y = f(y) - \langle u, \pi_{x,y}^{-1}(g_2) \rangle_x,$$

usi利用 $\pi_{x,y}$ 是等距映射这一事实。将两个不等式相加即可得出结果 t.

□

附录 C：度量空间中凸函数序列

引理 2 的证明

几乎必然收敛 由于 (M, d) 是一个适当的度量空间，因此它是可分的 [4, 命题 2.2。2]。设 M_0是 M的一个可数稠密子集。根据假设， $F_n(x) \xrightarrow[n\to\infty]{} F(x)$ 几乎必然成立，对所有 $x \in M_0$均成立。因此，以概率 1，收敛对所有 $x \in M_0$均成立，因为 M_0 是可数的。因此，由于 M_0 也稠密，定理 1适用，从而得到所需结果。

依概率收敛 这里我们使用以下事实。

Lemma 10. *[17, 引理 2 (ii) p.67]* 设 $X, X_1, X_2, \ldots$ 为实值随机变量。则，X_n 依概率收敛到 X 当且仅当 $(X_n)_{n\geq 1}$ 的每个子序列自身都有一个子序列几乎必然收敛到 X。

设 K 是 M 的一个紧子集，并考虑 $(\sup_{x\in K} |F_n(x) - F(x)|)_{n\geq 1}$的一个子序列。不失一般性（可以对该子序列重新编号），我们实际上考虑整个序列 $(\sup_{x\in K} |F_n(x) - F(x)|)_{n\geq 1}$，并证明它存在一个几乎必然收敛到零的子序列。设 M_0 是 M的一个可数稠密子集。将 M_0 中的元素记为x_1, x_2, ...

根据假设， $F_n(x_1)$依概率收敛到 $F(x_1)$，因此，它有一个子序列，记作$F_{\phi_1(n)}(x_1)$，几乎必然收敛到 $F(x_1)$。

现在，同样根据假设，$F_{\phi_1(n)}(x_2)$依概率收敛到 $F(x_2)$，因此，它有一个子序列，记作$F_{\phi_1\circ\phi_2(n)}(x_2)$，几乎必然收敛到 $F(x_2)$。

通过反复执行这一过程，对于所有整数 $k \geq 1$，我们构造出一个递增函数 $\phi_k : \mathbb{N}^* \to \mathbb{N}^*$，使得 $F_{\phi_1\circ\ldots\circ\phi_k(n)}(x_k)$ 几乎必然收敛到 $F(x_k)$，当 $n \to \infty$ 时。

最后，令 $\psi : \mathbb{N}^* \to \mathbb{N}^*$ 为由 $\psi(n) = \phi_1 \bigcirc \ldots \bigcirc \phi_n(n)$ 定义的函数，对所有 $n \geq 1$ 均成立。根据构造，$F_{\psi(n)}(x_k)$ 对所有 $k \geq 1$ 均几乎必然收敛到 $F(x_k)$。因此，以概率 1，$F_{\psi(n)}(x)$ 对所有 $x \in M_0$ 均几乎必然收敛到 $F(x)$。由引理 1 可知，$\sup_{x\in K} |F_{\psi(n)}(x) - F(x)|$ 几乎必然收敛到零，由此完成证明。

APPENDIX 部分：亚历山德罗夫曲率简介

本节将给出 CAT 空间的精确定义。

D.1 模型空间

首先，我们介绍一族模型空间，后文将借此定义局部和全局曲率界。设 $\kappa \in \mathbb{R}$。

$\kappa = 0$：欧几里得平面。设 $M_0 = \mathbb{R}^2$，并赋予其欧几里得度量。该模型空间对应零曲率，是一个测地空间，其中测地线唯一，且由线段给出。

$\kappa > 0$：球面。设 $M_\kappa = \frac{1}{\sqrt{\kappa}}\mathbb{S}^2$：这是嵌入在 $\mathbb{R}^3$ 中的二维欧几里得球面，球心为 0，半径为 $1/\sqrt{\kappa}$，并赋予弧长度量：$d_\kappa(x,y) = \frac{1}{\sqrt{\kappa}} \arccos(\kappa x^\top y)$，对所有 $x, y \in M_\kappa$ 均成立。这是一个测地空间，其中除对径点外，测地线均唯一，且由大圆弧给出。这里，大圆是球面与任意一个经过原点的平面的交集。

$\kappa < 0$：双曲空间集合 $M_\kappa = \frac{1}{\sqrt{\kappa}}\mathbb{H}^2$，其中 $\mathbb{H}_2 = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_3 > 0, x_1^2 + x_2^2 - x_3^2 = -1\}$。度量定义为 $d_\kappa = \frac{1}{\sqrt{-\kappa}}\operatorname{arccosh}(-\kappa x, y)$，对所有 $x, y \in M_\kappa$ 均成立，其中 $x, y = x_1y_1 + x_2y_2 - x_3y_3$。这是一个测地空间，其中测地线始终唯一，且由 M_κ 与经过原点的平面的交线弧段给出。

对于 $\kappa \in \mathbb{R}$，令 D_κ 为模型空间 M_κ的直径，即 $D_\kappa = \begin{cases} \infty & \kappa \leq 0 \\ \frac{\pi}{\sqrt{\kappa}} & \kappa > 0 \text{ 如果。} \end{cases}$ 如果

D.2 曲率界

设(M, d)为一个测地空间。(M, d)的曲率（下界或上界）概念是通过将 M 中的三角形与模型空间中具有相同边长的三角形进行比较来定义的。

Definition 3. M 中的一个（测地线）三角形由 M 中的三个点（顶点）以及连接它们的三条测地线段（边）组成。

给定三个点 x、y、$z \in S$，我们不严格地将顶点为 x、y、z 的三角形记作 $\Delta(x, y, z)$，而不说明其各边所选取的测地线段（点与点之间的测地线不一定唯一，例如在球面上，任意两个对径点之间的测地线就不唯一）。三角形 $\Delta = \Delta(x, y, z)$ 的周长定义为 $\mathrm{per}(\Delta) = d(x, y) + d(y, z) + d(x, z)$。它不依赖于边的选择。

Definition 4. 设 $\kappa \in \mathbb{R}$ 和 Δ 是 M 中的一个三角形，其周长为 $< 2D_\kappa$。模型空间 M_κ 中对应于 Δ 的比较三角形是一个与 Δ 边长相同的三角形 $\bar{\Delta} \subseteq M_\kappa$，也就是说，如果 $\Delta = \Delta(x, y, z)$，

¯然后 $\Delta = \Delta(\bar{x}, \bar{y}, \bar{z})$，其中 $\bar{x}, \bar{y}, \bar{z}$ 是满足条件的 M_κ 中的任意点

$$\begin{cases} d(x,y) = d_\kappa(\bar{x}, \bar{y}) \\ d(y,z) = d_\kappa(\bar{y}, \bar{z}) \\ d(x,z) = d_\kappa(\bar{x}, \bar{z}). \end{cases}$$

注意，在 M_κ 中，比较三角形在刚性运动意义下总是唯一的。现在我们可以定义曲率界了。直观地说，如果(M,d)中的所有三角形（周长小于 $2D_\kappa$）都比模型空间 M_κ中的对应比较三角形更瘦（分别更胖），那么我们就说它的全局曲率从上方（分别从下方）被 κ 所界。

Definition 5. 设 $\kappa \in \mathbb{R}$。当且仅当对于所有满足 $\mathrm{per}(\Delta) < 2D_\kappa$ 的三角形 $\Delta \subseteq M$，以及所有 x、$y \in \Delta$、d，均有 $(x,y) \le d_\kappa(\bar{x}, \bar{y})$（分别地，$d(x,y) \le d_\kappa(\bar{x}, \bar{y})$），其中 $\bar{x}$ 和 $\bar{y}$ 是比较三角形 Δ 中与 x 和 y 相对应的点时，我们称 (M,d) 的全局曲率上界（分别地，下界）为 κ。此时，我们称 (M,d) 为 CAT(κ)（分别地，CAT$^+$(κ)）空间。

Lemma 11. *[45,命题 3.1]*
设 (M,d) 是某个 $\kappa \in \mathbb{R}$ 下的 CAT(κ) 空间，并假设 M 的直径为 $D < D_\kappa/2$。那么，对于所有 $x_0 \in M$，函数 $d(\cdot, x_0)$ 在 M 上是凸的。此外，令 $\alpha_{D,\kappa} = \frac{\sqrt{\kappa}D}{\tan(\sqrt{\kappa}D)}$ 若 $\kappa > 0$, $\alpha_{D,\kappa} = 1$ 否则。那么，函数 $\frac{1}{2}d^2(\cdot, x_0)$ 是 $\alpha_{D,\kappa}$-强凸的，也就是说，它满足

$$\frac{1}{2}d^2(\gamma(t), x_0) \le \frac{1-t}{2}d^2(x, x_0) + \frac{t}{2}d^2(y, x_0) - \frac{\alpha_{D,\kappa}t(1-t)}{4}d(x,y)^2,$$

对于所有 $x, y \in M$ 和 $\gamma \in \Gamma_{x,y}$。

D.3 CAT 空间的例子

本节给出几个 CAT 空间的例子。首先，下面的引理是 [15, 定理 1A.6] 与 [16, 定理 9.2.9]，的直接推论，它们给出了一整类作为 CAT 空间的黎曼流形。

Lemma 12. 设（M,g）为一个单连通黎曼流形。假设所有截面曲率都不大于某个 $\kappa \in \mathbb{R}$。那么，M 是 CAT(κ)。

我们给出两个在最优传输和矩阵分析中尤为重要的例子。

1. 令 M 为所有 $p \times p$实对称正定矩阵的集合。赋予 M 度量 $d(A,B) = \|\log(A^{-1/2}BA^{-1/2})\|_F$，其中 log 是矩阵对数，$\|\cdot\|_F$是弗罗贝尼乌斯范数。于是，$d$ 继承自一个黎曼度量，且（M,d）是 CAT(0) [9]。这个度量空间在矩阵几何均值的研究中尤为重要。对于任意$A, B \in M$，它们的几何均值定义为 $A\#B = A^{1/2}(A^{-1/2}BA^{-1/2})^{1/2}A^{1/2}$，它是从 A 到 B的唯一中点，即 $A\#B = \gamma(1/2)$，其中 γ 是连接 A和 B的唯一测地线。

2. 令 M 为所有 $p \times p$ 实对称正定矩阵的集合，其谱包含于 $[\lambda_0, \infty)$，其中 $\lambda_0 > 0$ 为某个值。赋予 M Bures-Wasserstein度量，其定义如下。对于 A、$B \in M$，令 $d(A, B) = \mathrm{W}_2(\mathrm{N}_p(0, A), \mathrm{N}_p(0, B))$，其中 W_2 为 Wasserstein距离，$\mathrm{N}_p(0, A)$（分别地，$\mathrm{N}_p(0, B)$）是均值为 0、协方差矩阵为 A（分别为 B）的 p 维高斯分布。那么，d 也继承自黎曼度量 [10]，且 (M, d) 是 CAT($3/(2\lambda_0)$) 空间，这由 [43, 命题 2] 得出。

以下给出更多示例。

- 度量树是 CAT(κ) 对任意 $\kappa \in \mathbb{R}$ （因为其所有三角形都是平坦的） [15, 第二节 I.1。15]
- 欧几里得空间中光滑凸体的表面是 CAT(κ)，其中 $\kappa > 0$ 是该凸体补集的逼近半径上界，这是 [48, 定理 3.2.9]的推论。
- 系统发育树空间 [14] 是 CAT(0)。