arXiv:2303.01144v1 [math.ST] 2 Mar 2023

度量空间中经验重心的集中性

Victor-Emmanuel Brunel[1] 和 Jordan Serres[2]

[1]CREST-ENSAE，Victor-Emmanuel.Brunel@ensae.fr[2]
CREST-ENSAE，jordan.serres@ensae.fr

2023年3月3日

摘要

重心（又称 Fr´echet 均值）于20世纪40年代被引入统计学，随后在形状统计学领域广为人知，并进一步应用于最优传输和矩阵分析。它们为非欧几里得几何中的线性平均提供了最自然的推广，而线性平均或许是数据科学中最基础、应用最广泛的工具。在各种设定下，重心的渐近性质（如大数定律和中心极限定理）已得到确立，但其非渐近行为仍未得到充分理解。在这项工作中，我们针对 Alexandrov意义下具有非正曲率的度量空间中的独立同分布随机变量的重心，证明了有限样本集中不等式（即 Hoeffding 不等式和 Bernstein 不等式的推广）。作为推论，我们还为一种随机在线算法提供 PAC保证，该算法用于计算非正曲率空间中有限个点的重心。我们还讨论了将结果推广到曲率可能为正的空间。

1 引言

统计学和机器学习越来越多地面临存在于非线性空间中的数据。例如，空间统计学（如方向数据）、计算机断层成像（如形状统计学中经过刚性变换后采集的商空间数据）、经济学（如最优传输，其中数据为离散测度）等。此外，编码为超高维向量的数据，其内在维数可能小得多；例如，这些数据可能位于欧几里得空间中的低维子流形上。在这种情况下，利用数据可能具有的非线性几何结构，可以有力地显著降低所研究问题的维数。尽管越来越多的算法被开发出来以处理此类数据 [31, 34, 40, 41], 但不确定性量化方面的理论仍然十分有限，尤其是在机器学习中普遍存在的非渐近情形下。本文针对数据点的重心证明统计学结果；重心是将线性平均推广到非线性几何背景下最自然的方式。具体而言，我们将次高斯随机变量的概念推广到度量空间，并在非正曲率条件下证明 Hoeffding 集中不等式和 Bernstein集中不等式的相应版本。最后，我们讨论将这些结果推广到可能具有正曲率的度量空间的情形。

若完备度量空间 (M, d) 对任意点对 $(x, y) \in M$，都存在满足以下条件的 $m \in M$，则称其为非正曲率空间（简称 NPC 空间）：

$$d(z,m)^2 \leq \frac{1}{2}(d(z,x)^2 + d(z,y)^2 - \frac{1}{2}d(x,y)^2), \tag{1}$$

对所有 $z \in M$ 均成立。众所周知，NPC 空间具有许多良好的正则性。具体而言，任意两点 $x, y \in M$之间都存在唯一的（速度恒定的）测地线 $\gamma_{x,y}$，即连续映射 $\gamma = \gamma_{x,y} : [0,1] \to M$，满足 $d(\gamma(s)\ \gamma(t)) = |s-t|d(x,y)$，对所有 $s, t \in [0,1]$ 均成立。测地线将欧几里得空间中的线段推广为从一点 x 到另一点 y的最短路径。此外，式 (1) 中的点 z 唯一，即 x 与 y的中点，也就是 $z = \gamma_{x,y}(1/2)$。最后，距离函数 d 关于两个变量联合测地凸；并且对于所有 $x_0 \in M$，函数 $\frac{1}{2}d(x_0, \cdot)^2$ 都是 1-强测地凸的。若函数 f沿任意测地线 $\gamma_{x,y}, x, y \in M$均为凸函数（分别为强凸函数），即对所有 $x, y \in M$ 和 $t \in [0,1]$，都有 $f(\gamma_{x,y}(t)) \leq (1-t)f(x) + tf(y)$，则称其为测地凸函数（分别称为强测地凸函数）。关于 NPC 空间的更多信息，读者可参见 [36, 14]；关于度量空间中的凸性，详细论述见 [7]。以下列举几个例子，NPC 空间包括：

- 欧几里得空间和希尔伯特空间；
- 双曲空间，以及更一般的卡当–阿达马流形，即处处具有非正截面曲率的单连通、完备黎曼流形 [36, 命题3。1];

- 度量树：这些树嵌入欧几里得平面中，其中任意两个点之间的距离是它们之间（唯一的）最短路径的欧几里得长度；
- 对称正定矩阵构成的空间 $\mathcal{S}_p$（矩阵大小为 p，其中 $p \geq 1$ 为任意固定整数）配备如下距离：对于任意 A、$B \in \mathcal{S}_p$，

$$d(A,B) = \|\log(A^{-1/2}BA^{-1/2})\|_{\mathsf{F}},$$

其中 $\|M\|_{\mathsf{F}} = \sqrt{\mathrm{Tr}(MM^\top)}$ $\forall M \in \mathbb{R}^{p\times p}$ 是弗¨罗贝尼乌斯范数。这一构造在研究矩阵几何均值时十分重要，下面我们将在第 3.3 节对此作出说明。关于该空间和矩阵几何均值的详细介绍，请参阅 [9] 及其中的参考文献。

将平均这一概念从线性空间推广到度量空间，一种自然的方式是借助重心的概念。给定 x_1、……、$x_n \in M$（$n \geq 1$），x_1、……、x_n 的重心是指 $\sum_{i=1}^n d(x_i,b)^2$ 的任意极小值点 $b \in M$。不难验证，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、……、x_n 的平均值。不过请注意，这在一般的赋范空间中并不成立。在 NPC 空间中，x_1、……、x_n 的重心是一个强凸优化问题的解，因此总是存在且唯一。更一般地，给定 (M,d) 上具有二阶矩的概率分布 μ，可以将 μ 的重心定义为 $\int_M d(x,b)^2\,\mathrm{d}\mu(x)$ 的唯一极小值点 $b \in M$。这里，称 μ 具有二阶矩，当且仅当函数 $d(x_0,\cdot)^2$ 关于 μ 可积，对某个（因而根据三角不等式，对所有）$x_0 \in M$均成立。容易验证，对于任意点对 x、$y \in M$以及任意 $t \in [0,1]$，$(1-t)\delta_x + t\delta_y$ 的重心（又称 x 和 y的加权重心，其权重分别为 $1-t$ 和 t）就是 $\gamma_{x,y}(t)$。例如，x 和 y 的重心就是 x 和 y的中点。概率分布的重心可以视为将期望这一概念推广到非线性空间。值得庆幸的是，定义 μ 的重心实际上只要求 μ具有一阶矩；其重心是 $\int_M (d(x,b)^2 - d(x,b_0)^2)\,\mathrm{d}\mu(x)$ 的极小值点 b，其中 $b_0 \in M$ 任意固定。容易验证，该极小值点不依赖于 b_0 的选取。重心还可以理解为使期望位移成本最小的点，其中位移成本以二次距离衡量。因此，过去 15 年来，重心在最优传输理论中日益受到重视，尤其是在经济学理论中的应用 [2]。

重心最初由 [21] 于20世纪40年代引入统计学，后来由 [25], 引入，当时它们更广为人知的名称是 Fr´echet 均值或卡尔歇均值。它们随后在形状统计学 [26] 和最优传输 [2, 18, 29, 17, 27, 5, 6] 领域得到普及。

另一种计算上简单得多的重心构造方法，定义如下。给定 n 个点 x_1、……、x_n，归纳重心定义为点 s_n，其中 $s_1 = x_1$、$s_2 = \gamma_{s_1x_2}(1/2)$，且对于 $k = 3$、……、n，$s_k = \gamma_{s_{k-1}x_k}(1/k)$。重心与归纳重心并不重合（除非 M 是欧几里得空间或希尔伯特空间），这是因为非线性空间中的重心不具有结合性。与重心相比，归纳重心的优点是：当点 x_1、……、x_n 以在线方式依次到达时，可以轻松更新。计算归纳重心不需要求解优化问题，只需知道 M 中的测地线。然而，这种构造方法的一个缺点是它依赖于 x_1、……、x_n 的顺序。正如我们将在下文看到的，在统计学背景下，这一缺点非常小，因为在随机框架中，随着 n 变大，s_n 会越来越接近 x_1、……、x_n 的重心，这种依赖性也会消失。此外，正如 [34], 所指出的，当 x_i 可逐个获得时，x_1、……、x_n 的归纳重心是在线近端梯度算法的输出，该算法用于近似它们的重心。事实上，对于 $i = 1$、……、n，令 $f_i = d(\cdot,x_i)^2$。那么，x_1、……、x_n 的重心使凸函数 $f = f_1 + \ldots + f_n$ 最小化；一个从 $y_1 = x_1$ 开始的在线近端梯度算法会在每一步 $t = 2$、……、n 输出使 $f_t(y) + \frac{1}{2\lambda_t} d(y,y_{t-1})^2$ 最小化的点 $y_t \in M$，其中 λ_t 是变化的步长，此处为 $\lambda_t = (2t)^{-1}$。

设 μ 是 (M,d) 上具有两个（当然，一个也足够）矩的概率测度，且 b^* 是其重心。设 $X_1,\ldots,X_n$ 是 M 中服从分布 μ 的独立同分布随机变量。这里，$n \geq 1$ 是样本量，且为固定值。我们关注基于 $X_1,\ldots,X_n$ 对 b^* 的估计。设 b_n 是它们的重心（称为经验重心），S_n 是它们的归纳重心。

重心的存在性和唯一性通常是很难解决的问题 [1, 38, 39]。在各种设定下，经验重心的渐近理论已经得到充分理解，尤其是大数定律 [42] 和黎曼流形中的中心极限定理（为推导中心极限定理，在 M 上假设存在光滑结构是很自然的）[12, 13, 11, 20, 19]。迄今为止，已证明的非渐近结果寥寥无几，而且大多需要相当复杂的条件。首先，[36, 定理 4.7] 给出了非正曲率空间中 $d(S_n,b^*)^2$ 期望值的上界。具体而言，$\mathbb{E}[d(S_n,b^*)^2] \leq \frac{\sigma^2}{n}$其中，$\sigma^2 = \mathbb{E}[d(X_1,b^*)^2]$ 可解释为 X_1的方差。[28, 推论 11] 在额外假设 (M,d) 的曲率下界有界的条件下，为 b_n给出了同样的不等式。从较高层面来看，这意味着空间 (M,d) 不存在分支（即测地线不会分叉，这与度量树等情形不同），并由此确保 M的切锥具有一定的正则性。他们还将结果推广到了不一定是非正曲率空间、但满足所谓拥抱条件的空间 (M,d)。然而，除非是非正曲率空间，目前还没有明确的度量空间满足

这样的条件。对于经验重心和归纳重心，也已有一些非渐近的高概率界。 [28] 提出了次高斯随机变量的定义（式（3.10））， 与我们下面给出的定义密切相关，并证明（定理 12）， 在上述拥抱条件下，独立同分布次高斯随机变量的经验重心围绕总体重心 n具有近似次高斯尾界（带有一个随 b*指数速度衰减的余项）。 [3] 在一些关于（M， d）的度量熵条件下，为独立同分布有界随机变量的经验重心 b_n 建立了具有非参数速率的集中不等式；其中一些条件在精神上类似于要求 M 具有有限维数。最后，与我们工作最为密切相关的是，[22] 证明了非正曲率空间中独立同分布有界随机变量的归纳重心满足 Hoeffding 型不等式，尤其关注度量树和有限维 Hadamard流形。下文的推论 3.4 从两个方面改进了他的结果。首先，我们显著改进了常数。其次，我们的上界包含两项。第一项是偏差项，不包含概率水平 δ。只有这一项可能涉及空间的维数，具体通过 σ^2 的方差体现。 X_1因此，第二项即次高斯项（因为它与 $\sqrt{\log(1/\delta)}$成正比）与 X_1的方差相互独立。

本文组织如下。在第2节中，我们介绍测度集中理论中的工具，以便推导经验重心收敛的非渐近界。在第3节中，我们为 NPC 空间中的经验重心推导 Hoeffding 和 Bernstein 集中不等式。在第4节中，我们将前述结果推广到随机变量不满足同分布的情形（但其理论重心相同）。最后，在第5节中，我们讨论可能具有正曲率的空间，并推导一个依赖于空间半径的最优速率集中界。

下文中，对于任意正整数 n 以及任意度量空间 (M,d)，我们用 d_1 表示 M^n 上的 L^1 乘积度量（为简便起见，不特别注明其对 n 的依赖），定义为 $d_1((x_1,\dots,x_n),(y_1,\dots,y_n)) = d(x_1,y_1) + \dots + d(x_n,y_n)$。

2 度量空间中的拉普拉斯变换与次高斯随机变量

2.1 拉普拉斯变换

本节汇集了有关度量空间上概率测度的拉普拉斯变换的信息。这将使我们能够精确研究非正曲率空间中重心的集中现象，尤其是处理此类空间中的次高斯随机变量。设 (M,d) 为一个度量空间（不一定是 NPC 空间）。记 $\mathcal{F}$为所有满足 1-Lipschitz 条件的函数 $f\colon M\to\mathbb{R}$ 的类，即对于所有 $x,y\in M$，均有 $|f(x)-f(y)|\le d(x,y)$。设 X为 M中的一个随机变量。对于所有 $k\ge 1$，若 $\mathbb{E}[d(X,x_0)^k]$有限，则称 X具有 k矩，其中 x_0 是 M 中的任意点（注意，此定义不依赖于 x_0的选取）。至少具有一个矩的随机变量 X 的拉普拉斯变换定义如下（参见 [30, 第 1.6] 节）

$$\Lambda_X(\lambda) := \sup_{f\in\mathcal{F}} \mathbb{E}[e^{\lambda(f(X)-\mathbb{E}[f(X)])}], \quad \lambda\ge 0. \tag{2}$$

注意，若 X 有一个矩，则对所有 $f\in\mathcal{F}$， $f(X)$ 也有一个矩。我们强调拉普拉斯变换的以下性质，其证明可见 [30, 命题 1.15]。

引理 2.1. 若 X_1，……， X_n 是定义在 (M,d) 上且至少有一个矩的独立随机变量，则拉普拉斯变换 在 (M^n,d_1) 上的随机向量 $(X_1,\dots,X_n)$ 满足

$$\Lambda_{(X_1,\dots,X_n)} \le \Lambda_{X_1}\dots\Lambda_{X_n},$$

这里回顾一下， d_1 是 M^n 上的 L^1 乘积度量。

最后，我们陈述拉普拉斯变换的一项性质：对随机变量应用 Lipschitz函数时，该性质成立。

引理 2.2. 设 $(M^{(1)},d^{(1)})$ 和 $(M^{(2)},d^{(2)})$ 为度量空间，且 $\Phi\colon M_{(1)}\to M_{(2)}$是一个 L-Lipschitz 函数，其中 $L>0$。那么，对于所有至少具有一个矩的随机变量 X in $M^{(1)}$ ，

$$\Lambda_{\Phi(X)}(\lambda) \le \Lambda_X(\lambda L), \quad \forall\lambda\ge 0.$$

2.2 次高斯随机变量

定义 2.3。若随机变量 X 属于 (M,d)，当且仅当其拉普拉斯变换满足以下条件时，称其为 K^2-次高斯（$K\ge 0$）

$$\Lambda_X(\lambda) \le e^{\frac{\lambda^2K^2}{2}}, \quad \forall\lambda\ge 0. \tag{3}$$

次高斯随机变量的这一定义比欧几里得空间中的标准定义更强。后者的拉普拉斯变换仅定义为对线性函数取上确界，而非对所有 Lipschitz 函数取上确界。不过，在度量空间的背景下，由于缺少线性函数，这一定义似乎最为恰当；而且，正如下面的引理 2.7 将要说明的，它还允许我们处理有界随机变量。此外，与 [28], 中提出的定义不同，这一定义不依赖于 M中的任何参考点。

次高斯随机变量已有充分研究，在欧几里得空间中发挥着非常重要的作用。特别是，已知它们具有良好的集中性质 [37, 第 2 节5]。注意，次高斯随机变量自动具有无限多个矩。

下　　述引理推广了度量空间中次高斯随机变量的重要性质　　.

L引理 2.4。设 X 是 (M,d) 中的一个随机变量。以下命题等价：　　$t.$

(i) X 是 K^2-次高斯（$K>0$）。

(ii) 对所有 $f\in\mathcal{F}$,

$$P(|f(X)-\mathbb{E}[f(X)]|\geq t)\leq 2e^{-t^2/(2K^2)}.$$

(iii) 对所有 $f\in\mathcal{F}$,

$$\mathbb{E}[e^{a(f(X)-\mathbb{E}[f(X)])^2}]\leq 2,$$

其中 $a=\log\frac{2}{16eK^2}$。

特别是，引理 2.4 表明，如果 X 是 K^2-次高斯的，其中 $K>0$，那么对于任意 $\delta\in(0,1)$ 以及任意固定点 $x_0\in M$，以下结论以至少 $1-\delta$ 的概率成立

$$d(X,x_0)\leq\mathbb{E}[d(X,x_0)]+K\sqrt{2\log(2/\delta)}.$$

接下来的两个 命题表明，次高斯性质在乘积以及 Lipschi 变换下得以保持　　tz 变换。

命题2.5. 设 $X_1,\ldots,X_n$ 是 M 中的独立随机变量，并且对于某个 $K_i>0$，每个 X_i 都是 K_i^2-次高斯的。那么，　n 元组 $(X_1,\ldots,X_n)$ 在乘积度量空间 (M^n,d_1) 上是 $(K_1^2+\ldots+K_n^2)$-次高斯的。

证明。这是引理2.1的直接推论。 □

命题2.6. 设 $(M^{(1)},d^{(1)})$ 和 $(M^{(2)},d^{(2)})$ 是度量空间，且 X 是 M_1 中的一个随机变量。令 $K,L>0$。如果 X 是 K^2-次高斯的，且 $\Phi\colon M^{(1)}\to M^{(2)}$ 是 L-Lipschitz 的，那么 $\Phi(X)$ 是 (L^2K^2)-次高斯的。

证明。设 $g\colon M^{(2)}\to\mathbb{R}$ 是一个 1-Lipschitz 函数，使得 $g(\Phi(X))$ 是中心化的，并令 $\tilde{g}:=L^{-1}g\circ\Phi$。那么，$\tilde{g}\colon M^{(1)}\to\mathbb{R}$ 是 1-Lipschitz 的，且 $\tilde{g}(X)$ 是中心化的，因此有

$$\begin{aligned}\mathbb{E}e^{\lambda g(\Phi(X))-\mathbb{E}[g(\Phi(X))]}&=\mathbb{E}e^{\lambda L(\tilde{g}(X)-\mathbb{E}[\tilde{g}(X)])}\\&\leq\Lambda_\mu(\lambda L)\\&\leq e^{\lambda^2L^2K^2/2}.\end{aligned}$$

最后，我们对所有此类函数 g 取上确界。 □

最后，我们陈述一个来自 [30], 的重要引理。与针对实值随机变量的 Hoeffding引理类似，该引理指出，有界随机变量总是次高斯的。

引理2.7。 *[3　　0，命题1。16]* 设 X 是度量空间 $(M,\ d)$中的有界随机变量，
即，$d(x_0,X)$ 对于某些必然成立。和 $C>0$。那么，　X 是 $4C^2$-次高斯的。

3 非正曲率空间中经验重心的集中性

3.1 重心的Lipschitz性质

要证明非正曲率空间中经验重心和归纳重心的集中性，最后一个要素是它们的Lipschitz性质，我们将在下一个命题中陈述这一性质。记 $B_n\colon\ M^n\to M$ 为将任意 n-元组映射到其（唯一确定的）重心的函数˜，并记 B_n 为将任意 n-元组映射到其归纳重心的函数。

˜命题 3.1. 假设 (M,d) 是 NPC 空间。那么，函数 B_n 和 B_n 都是 $(1/n)$-Lipschitz 的，M^n 配备 L^1 乘积度量 d_1。

˜证明。 归纳重心函数的证明 B_n 可由简单的归纳法得出，详见 [22, 引理 3。1]。

B_n 的 1-Lipschitz性质可由两种不同的论证得出；我们在此一并给出，因为二者都颇具启发性。令 $\mathcal{P}^1(M)$ 表示 (M,d) 上所有具有有限一阶矩的概率测度构成的集合。对于 $\mu \in \mathcal{P}^1(M)$，令 $B(\mu)$ 为其重心，即 $\mathbb{E}[d(X, b)^2 - d(X,b_0)^2]$, 的（唯一）极小值点 $b \in M$，其中$X \sim \mu$，且 $x_0 \in M$ 任意固定。

第一种论证源自 [31, 定理 3.4]，，该定理揭示了重心与其归纳版本之间的确定性联系。设 $(x_1,\ldots,x_n)$ 和 $(y_1,\ldots,y_n)$ 是 M 中的两个 n元组。我们将这两个 n元组扩展为周期无限序列：对于所有正整数 k，$x_k = x_{r_k}$ 和 $y_k = y_{r_k}$，令r_k 为 1 与 n 之间唯一的整数，使得 $k - r_k$ 是 n的倍数。于是，[31, 定理 3.4] 表明，$B_k(x_1,\ldots,x_k)$ 和 $B_k(y_1,\ldots,y_k)$ 分别收敛到 $B_n(x_1,\ldots,x_n)$ 和 $B_n(y_1,\ldots,y_n)$，当 $k \to \infty$时。˜此外，得益于上文证明的 B_{qn} 的 $(1/(qn))$-Lipschitz特性，对于所有正整数 q，有

$$d(\tilde{B}_{qn}(x_1,\ldots,x_{qn}), \tilde{B}_{qn}(y_1,\ldots,y_{qn})) \leq \frac{1}{qn}\sum_{k=1}^{qn} d(x_k,y_k) = \frac{1}{n}\sum_{k=1}^{n} d(x_k,y_k).$$

令 $q \to \infty$ 趋于极限，可得 B_n 的 $(1/n)$-Lipschitz性质。

第二个论证使用詹森不等式，这意味着重心泛函 B 在赋予 Wasserstein距离 W_1 [36, 的 $\mathcal{P}^1(M)$ 上是压缩的，定理 6.3]。更确切地说，对于所有概率测度 $\mu,\nu \in \mathcal{P}^1(M)$，都有 $d(B(\mu)\ B(\nu)) \leq W_1(\mu,\nu)$，其中 $W_1(\mu,\nu) = \inf_{X\sim\mu, Y\sim\nu} \mathbb{E}[d(X,Y)]$。现在，固定 n-元组 $(x_1,\ldots,x_n)$ 和 $(y_1,\ldots,y_n)$，它们属于 M^n，并令 $\mu = n^{-1}\sum_{i=1}^n \delta_{x_i}$ 和 $\nu = n^{-1}\sum_{i=1}^n \delta_{y_i}$。显然， $B(\mu) = B_n(x_1,\ldots,x_n)$，且 $B(\nu) = B_n(y_1,\ldots,y_n)$。此外， $W_1(\mu,\nu) \leq \frac{1}{n}(d(x_1,y_1) + \ldots + d(x_n,y_n))$；这一点可通过取 μ 和 ν 的耦合 (X,Y) 来看出，使得 $P(X = a_i, Y = b_i) = \frac{1}{n}, i = 1,\ldots,n$。 □

备注 3.2。上述命题实际上只要求 (M,d) 是布斯曼空间，即一个度量空间 (M,d)，使得度量 d 关于两个变量联合地测地凸。这一性质弱于非正曲率性质，因为任何 NPC 空间都是布斯曼空间；但例如，严格凸赋范空间是布斯曼空间，却不是 NPC 空间。

3.2 次高斯随机变量重心的集中不等式

现在我们可以陈述第一个主要结果，该结果蕴含了 NPC 度量空间中独立同分布次高斯随机变量的经验重心和归纳重心的集中性。该结果直接由上述命题2.5 和 2.6 得到。

定理 3.3。设 (M,d) 是一个 NPC 空间，且 $X_1, ..., X_n$ 是 (M,d) 中的独立随机变量，使得对所有 $i = 1,\ldots,n$， X_i 是 K_i^2-次高斯的，其中 $K_i > 0$ 为某个值。则 $X_1, ..., X_n$ 的经验重心和归纳重心都是 $\frac{K_1^2+\ldots+K_n^2}{n^2}$-次高斯的。

因此，根据引理2.4，对于所有 1-Lipschitz函数 $f:\ M \to \mathbb{R}$ 以及所有 $\delta \in (0,\ 1)$，以至少 $1-\delta$ 的概率，有

$$f(T_n) \leq \mathbb{E}[f(T_n)] + \bar{K}_n\sqrt{\frac{\log(1/\delta)}{n}}, \tag{4}$$

其中 T_n i s 是 $X_1,\ldots,X_n$ 以及 $\bar{K}_n = \sqrt{K_1^2+\ldots+K_n^2}$的经验重心或归纳重心 .

假设 $X_1, \ldots, X_n$ 是独立同分布的，并令 b^* 为它们的（总体）重心。记 $\hat{b}_n$ 为它们的经验 重心， b_n 为它们的归纳重心。特别值得关注的是，在式 (4) 中取 $f = d(\cdot, b^*)$。令 $\sigma^2 = \mathbb{E}[d(X_1,b^*)^2]$ 表示 X_1的方差。当 T_n 为 $X_1,\ldots,X_n$的归纳重心时，由 Sturm 大数定律 [36, 定理 4.7] 可得 $\mathbb{E}[d(T_n,b^*)] \leq \frac{\sigma}{\sqrt{n}}$。当 T_n 为 $X_1,\ldots,X_n$的经验重心时，若空间 (M,d) 还满足曲率下界有界，则根据 [28, 推论 11] 可得同一不等式。因此，我们得到以下推论。

推论 3.4。 在与定理 3.3相同的假设和记号下，令 T_n 为 X_1, ..., X_n的经验ˆ重心或归纳重心。若 T_n 是经验重心 b_n，则进一步假设 (M,d) 的曲率下界有界。那么，对所有 $\delta \in (0,1)$，以下结论以至少 $1-\delta$ 的概率成立:

$$d(T_n,b^*) \leq \frac{\sigma}{\sqrt{n}} + K\sqrt{\frac{\log(1/\delta)}{n}}$$

现在我们可以在全局NPC空间中陈述Hoeffding不等式。

推论 3.5. 设 (M,d) 是一个 NPC 空间，且 $X_1,...,X_n$ 是独立同分布随机变量，并假设几乎必然有 $X_1 \in B(x_0,C)$，其中 $x_0 \in M$ 和 $C>0$ 为某些常数。设 T_n 是 $X_1,\dots,X_n$ 的经验重心或归纳重心。若 T_n 是经验重心 b_n，则进一步假设 (M,d) 的曲率下界有界。那么，对于所有 $\delta \in (0,1)$，以下结论以至少 $1-\delta$ 的概率成立：

$$d(T_n,b^*) \le \frac{\sigma}{\sqrt{n}} + 2C\sqrt{\frac{\log(1/\delta)}{n}}.$$

该集中不等式的左侧包含两项。第一项是偏差项，仅用于控制从 T_n 到 b^* 的期望距离；第二项是随机项，其中包含概率水平 δ。如果 (M,d) 是希尔伯特空间（它是非正曲率空间的一个特例），那么该不等式实际上可写为

$$d(T_n,b^*) \le \sqrt{\frac{\mathrm{tr}\Sigma}{n}} + 2C\sqrt{\frac{\log(1/\delta)}{n}},$$

其中 Σ 是 X_1的协方差算子。例如，若 $M=\mathbb{R}^p$ 配备欧几里得结构，且Σ 是单位矩阵，那么 $\mathrm{tr}\Sigma = p$，并且 M 的维数只出现在偏差项中，而不出现在随机项中。这就是推论3.4 显著优于 Funano 的结果的原因 [22]。

作为推论 3.4 的结果，我们得到以下算法 PAC 保证，用于计算 NPC 空间 (M,d) 中 n 个固定点 $x_1,\dots,x_n$ $(n\ge 1)$ 的重心。固定 $\varepsilon>0$ 和 $\delta\in(0,1)$。从 1 到 n 独立、均匀随机地抽取 m 个整数 $I_1,\dots,I_m$，并令 $X_1=x_{I_1},\dots,X_m=x_{I_m}$，它们是分布为 $\mu = n^{-1}\sum_{i=1}^n \delta_{x_i}$ 的独立同分布随机变量。最后，迭代计算 $\tilde{b}_m$，即 $X_1,\dots,X_m$ 的归纳重心。

推论 3.6。记 b^* 为 $x_1,\dots,x_n$的重心。令 D 为集合 $\{x_1,\dots,x_n\}$的直径。那么，如果 $m \gtrsim \frac{D^2}{\varepsilon^2}\max(1,\log(1/\delta))$，则 $d(\tilde{b}_m,b^*) \le \varepsilon$ 的概率至少为 $1-\delta$。

该推论的证明只用到了这样一个事实：方差 σ^2 属于 μ，其支撑集的直径为 D，其上界为 $4D^2$。据我们所知，推论3.6 是首个针对度量空间中重心计算的 PAC 保证。请注意，利用下文的定理3.7 ，还可以进一步改进这一界，参见第3.4节。

与这一保证相比，[31, 定理 3.4] 为寻找 $x_1,\dots,x_n$的重心的 ε-近似提供了确定性保证，所需步骤为 $n^{(D^2+\sigma^2)}$ ε^2：他们的算法复杂度比我们的高 n 倍，~~其中~~ n 为输入点的数量。

或许令人惊讶的是，推论3.6 给出的样本复杂度上界并不需要对空间的某种维数概念作出任何限制 M。然而，该算法隐含地要求在归纳重心序列的每次迭代中计算~ ~测地线段 b_1，……，b_m。这类计算的复杂度通常取决于 M的维数（在某种意义上）M， d。当(M,d)是度量树时，计算归纳重心只需确定任意两点之间的最短路径，这可以高效完成。

3.3 一个例子：矩阵几何均值

本节考虑 n 对称正定矩阵A_1，……， $A_n \in \mathcal{S}_p(p\ge 1)$的几何均值的计算。回顾一下， A_1，……， A_p的几何均值是它们相对于度量 $d(A,B) = \|\log(A^{-1/2}BA^{-1/2}\|_F$的重心，对所有 $A,B\in\mathcal{S}_p$均成立。已知配备这一度量的 $\mathcal{S}_p$是一个NPC 空间 [9, 5] $A\ B\in\mathcal{S}_p$ $A\quad B$ $A\quad B$ $=$命题。两个矩阵 , 的几何均值是矩阵 s $A^{1/2}\ A^{-1/2}BA^{-1/2\ 1/2}A^{1/2}$ # ( ); 更一般地，和之间的测地线段由 $\gamma_{A\ B}\quad =,(\ )(\ )$ 给出，也记作 # ，对所有 均 $A^{1/2}\ A^{-1/2}BA^{-1/2\ s}A^{1/2}$ $A\quad {}_sB$ $s\in[0,1]$ 成立。

归纳重心序列定义如下：

- $S_1 = A_1$;
- $S_2 = S_1 \# A_2 = A_1^{1/2}(A_1^{-1/2}A_2A_1^{-1/2})^{1/2}A_1^{1/2}$;
- For $k=2,\dots,m$, $S_k = S_{k-1}\#_{\frac{1}{k}}A_k = S_{k-1}^{1/2}(S_{k-1}^{-1/2}A_kS_{k-1}^{-1/2})^{1/k}S_{k-1}^{1/2}$.

精确计算每个 S_k、 $k=2$、……、 m，都需要进行矩阵乘积和特征分解，其复杂度取决于矩阵的大小 p。事实上，有更快的方法可以计算 $A\#_sB$ 的良好近似值，对于 A、 $B\in\mathcal{S}_p$ 和 $s\in[0,1]$，例如使用积分表示和高斯求积：例如，详情请参阅 [8, 35] 。

3.4 伯恩斯坦不等式：小方差情况下的改进

本节中，我们针对 X_i 的方差较小（即远小于其支撑集的直径）的情况，推导一个改进 Hoeffding引理的伯恩斯坦不等式。

定理 3.7. 设 (M,d) 是一个 NPC 空间，且 $X_1,\dots,X_n$ 是 (M,d) 中的独立同分布随机变量。假设对于某些 $x_0\in M$ 和 $C>0$，$d(X_1,x_0)\le C$ 几乎必然成立。那么，对于所有 1-Lipschitz函数 $f: M\to\mathbb{R}$，且对于所有 $\delta\in(0,1)$，

$$f(T_n)\le \mathbb{E}[f(T_n)]+\min\left(2\sigma\sqrt{\frac{\log(1/\delta)}{n}},\frac{8C\log(1/\delta)}{3n}\right),$$

其中，T_n 表示 X_1,...,X_n 的经验重心或归纳重心，而 $\sigma^2=\mathbb{E}[d(X_1,b^*)^2]$ 表示 X_1 的方差，其中 b^* 是 X_1 的分布重心。

当 X_i 的方差非常小时，即显著小于其支撑集的直径 C 时，这个不等式大大改进了推论 3.4。取 $f=d(\cdot,b^*)$，我们得到非正曲率空间中的 Bernstein不等式的以下形式。

推论 3.8. 在与定理 3.7相同的假设和记号下，设 T_n 为 ˆ经验重心或归纳重心 $X_1,\dots,X_n$。若 T_n 为经验重心 b_n，则进一步假设 (M,d) 的曲率有下界。那么，以至少 $1-\delta$的概率，有

$$d(T_n,b^*)\le\frac{\sigma}{\sqrt{n}}+\max\left(2\sigma\sqrt{\frac{\log(1/\delta)}{n}},\frac{8C\log(1/\delta)}{3n}\right).$$

定理 3.7的证明。对于所有 $\lambda\ge 0$和 $f\in\mathcal{F}$，令 $\psi(\lambda,f)=\log\mathbb{E}[e^{\lambda(f(X_1)-\mathbb{E}[f(X_1)])}]$,对于 $\lambda\ge 0$和 $f\in\mathcal{F}$。利用不等式 $\log(u)\le 1-u$，对于所有 $u>0$，有

$$\psi(\lambda,f)\le \mathbb{E}\left[e^{\lambda(f(X_1)-\mathbb{E}[f(X_1)])}-1-\lambda(f(X_1)-\mathbb{E}f(X_1))\right],$$

对于所有 $\lambda\ge 0$和 $f\in\mathcal{F}$。由于 $X_1\in B(x_0,C)$ 几乎必然成立，我们有 $f(X_1)-\mathbb{E}f(X_1)\le 2C$，对于所有 $f\in\mathcal{F}$。因此，由于映射 $u\mapsto\frac{e^u-1-u}{u^2}$ 是递增的，我们得到

$$\begin{aligned}\psi(\lambda,f)&\le\mathbb{E}\left[\frac{e^{2\lambda C}-1-2\lambda C}{4C^2}(f(X_1)-\mathbb{E}f(X_1))^2\right]\\&=\frac{e^{2\lambda C}-1-2\lambda C}{4C^2}\mathrm{Var}(f(X_1)).\end{aligned}$$

此外，对于所有 $f\in\mathcal{F}$，$\mathrm{Var}(f(X_1))\le\sigma^2$。事实上，

$$\begin{aligned}\mathrm{Var}(f(X_1))&=\mathbb{E}\left[(f(X_1)-\mathbb{E}[f(X_1)])^2\right]\\&\le\mathbb{E}\left[(f(X_1)-f(b^*))^2\right]\\&\le\mathbb{E}\left[d(X_1,b^*)^2\right]\\&=\sigma^2,\end{aligned}\tag{5}$$

其中（5）由以下事实可得：$\mathbb{E}\left[\left(f\left(X_1\right)-\mathbb{E}[f\left(X_1\right)]\right)^2\right]\le\mathbb{E}\left[\left(f\left(X_1\right)-a\right)^2\right]$对于所有 $a\in\mathbb{R}$，尤其对于 $a=f\left(b^*\right)$。

由引理2.1可知:

$$\Lambda_{X_1,\dots,X_n}(\lambda)\le e^{\frac{n\sigma^2}{4C^2}(e^{2\lambda C}-1-2\lambda C)},$$

对于所有 $\lambda\ge 0$。现在，令 T_n 为 X_1，.....，X_n的经验重心或归纳重心。由命题3.1 和引理2.2 可得:

$$\Lambda_{T_n}(\lambda)\le e^{\frac{n\sigma^2}{4C^2}(e^{2\lambda C/n}-1-2\lambda C/n)},$$

对于所有 $\lambda\ge 0$。

根据切尔诺夫界，对于所有 $f\in\mathcal{F}$，我们得到:

$$\begin{aligned}&P\left(f(T_n)-\mathbb{E}f(T_n)\ge t\right)\\&\le e^{-\lambda t}\mathbb{E}e^{\lambda(f(T_n)-\mathbb{E}f(T_n))}\\&\le e^{-\lambda t+\log\Lambda_{T_n}(\lambda)}\\&\le\exp\left[-\lambda t+\left(e^{\frac{2\lambda C}{n}}-1-\frac{2\lambda C}{n}\right)\frac{n\sigma^2}{4C^2}\right].\end{aligned}$$

通过对 $\lambda \geq 0$ 进行优化，我们发现最优值为 $\lambda = \frac{n}{2C}\log(1+\frac{2Ct}{\sigma^2})$，因此，

$$P\left(f(T_n) - \mathbb{E}f(T_n) \geq t\right) \leq \exp\left(-\frac{n\sigma^2}{4C^2}\, h\left(\frac{2Ct}{\sigma^2}\right)\right)$$

其中 $h(u) = (1+u)\log(1+u) - u$，对所有 $u \geq 0$ 均成立。由于对所有 $u \geq 0$，$h(u) \geq {}_2{}^{u^2/(1+\frac{u}{3})}$，我们得到

$$P\left(f(T_n) - \mathbb{E}f(T_n) \geq t\right) \leq \exp\left(-\frac{nt^2}{2(\sigma^2 + \frac{2Ct}{3})}\right),$$

由此得到所需结果。 □

注 3.9。在上述集中不等式中，当考虑经验重心时，我们假设空间 (M,d) 的曲率有下界。这样做是为了能够应用 [28, 推论 11]中证明的 $d(b_n, b^*)$ 期望上界。事实上，对于一般的 NPC 空间，目前尚无 $\mathbb{E}[d(b_n,b^*)]$ 的上界（不过，S_n [36, 定理 4.7] 中的上界适用于任意 NPC 空间）。我们确实认为，$\mathbb{E}[d(\hat{b}_n,b^*)^2] \leq \frac{\sigma^2}{n}$ 的上界在所有 NPC 空间中都成立，包括那些曲率无下界的空间，也就是说，包括可能发生分支的空间。没有曲率下界的 NPC 空间中，最简单也最自然的例子是度量树。事实上，对于所有 $\kappa \in \mathbb{R}$，此类空间都是 CAT(κ)，因为所有测地三角形都是平坦的（形式化定义见第 5.1 节）。在度量树中，重心具有黏性，即使对某些点 $x_1, \ldots\ldots, x_n$ 作出较大扰动，也可能不会改变其重心 [36, 命题 5.7]。尤其是，这类空间中的重心更加刚性，而且已知经验重心具有更快的渐近速率 [24]。因此，我们预计它们的表现不会比曲率有下界的 NPC 空间更差。

定理 3.7 推出以下算法推论，它是对推论 3.6的改进。

推论 3.10。 回顾推论 3.6中的记号。令 D 为集合 $\{x_1,\ldots,x_n\}$ 的直径，且$\sigma^2 = \min_{b\in M}\frac{1}{n}\sum_{i=1}^n d(x_i,b)^2$。那么，若

$$m \gtrsim \max\left(\frac{\sigma^2}{\varepsilon^2}, \frac{D}{\varepsilon}\right)\log(1/\delta), \tag{6}$$

~以至少 $1-\delta$ 的概率，有 $d(b_m,b^*) \leq \varepsilon$。

实际上，计算 σ^2 当然可能代价高昂。不过，在（6）中，可以用任何上界代替它，例如，对于任意固定的 $b \in M$，可以使用 $\frac{1}{n}d(x_i,b)^2$，或者使用 $\frac{1}{n^2}\sum_{1\leq i,\ j\leq n} d(x_i,x_j)^2$。计算 $\frac{1}{n}d(x_i,b)^2$ 只需要n 次计算，但 b 的选择可能并非最优。另一方面，计算 $\frac{1}{n^2}\sum_{1\leq i,\ j\leq n} d(x_i,x_j)^2$需要二次（关于 n）次计算，但能给出关于 σ^2的紧上界，误差至多为 2 倍。事实上，有如下引理。

引理 3.11. 设 X 是 M 中具有两个矩的随机变量，且 Y 是 M 的独立副本。设 $\sigma^2 = \min_{b\in M}\mathbb{E}[d(X,b)^2]$。那么，

$$\sigma^2 \leq \mathbb{E}[d(X,Y)^2] \leq 2\sigma^2.$$

证明。根据三角不等式，右侧显然成立（且不要求 X 和 Y 独立）。对于左侧，记 $F(b) = \mathbb{E}[d(X,b)^2]$, 对所有 $b \in M$ 均成立。那么，$\sigma^2 = \min_{b\in M} F(b)$，且由 X 和 Y 独立可得 $\sigma^2 \leq F(Y) = \mathbb{E}[d(X,Y)^2|Y]$。对两边取期望即可得到结论。 □

定理中给出的集中不等式 3.4 和 3.8 可扩展到以下情形： X_i 相互独立但不同分布；下一节将讨论这一问题。

4 非同分布随机变量的集中不等式

在本节中，我们将推论 3.4 和 3.8 推广到 $X_1,\ldots,X_n$ 相互独立但不同分布，且具有相同的重心 b^*这一情形。

定理 4.1。设 (M,d) 是一个 NPC 空间，且 $X_1,\ldots,X_n$ 是具有相同重心 $b^* \in M$ 的独立随机变量。进一步假设，对所有 $i = 1,\ldots,n$, $X_i \in B(x_i,C_i)$ 几乎必然成立，其中某个 $x_i \in M$ 和

$\hat{C}_i > 0$。设 T_n 是 $X_1, \ldots, X_n$ 的经验重心或归纳重心。若 $T_n = b_n$，进一步假设 M 具有曲率下界。那么，对所有 $\delta \in (0, 1)$，以下结论以至少 $1 - \delta$ 的概率成立：

$$d(T_n, b^*) \leq \frac{\overline{\sigma}_n}{\sqrt{n}} + \overline{C}_n \sqrt{\frac{\log(1/\delta)}{n}},$$

其中 $\overline{\sigma}_n = \sqrt{\frac{\sigma_1^2 + \ldots + \sigma_n^2}{n}}$和 $\overline{C}_n = \sqrt{\frac{C_1^2 + \ldots + C_n^2}{n}}$.

证明。证明与命题3.4的证明类似，主要区别在于如何界定偏差项$\mathbb{E}[d(T_n, b^*)]$。当 T_n 是经验重心时，仔细考察 [28, 推论 11] 的证明可知，$\hat{}$ X_i不必同分布，并且可以直接得到 $\mathbb{E}[d(b_n, b^*)] \leq \frac{\overline{\sigma}_n}{\sqrt{n}}$。当 T_n 是归纳重心时，我们调整 [36, 定理 4.7] 的证明，得到以下引理。

□

引理 4.2. 设 $X_1, \ldots, X_n$ 是 NPC 空间 (M, d) 中具有二阶矩且重心相同的独立随机点。那么，

$$\mathbb{E}[d(S_n, b^*)^2] \leq \frac{\sigma_1^2 + \ldots + \sigma_n^2}{n^2},$$

其中 $\sigma^2 = \mathbb{E}[d(X_i, b^*)^2]$ 是 X_i 的方差，适用于 $i = 1, \ldots, n$。

证明。该引理的证明与 [36, 定理 4.7] 的证明思路相同，并对 n进行归纳。记 $V_n = \mathbb{E}[d(S_n, b^*)^2]$，并记 $F_n(x) = \mathbb{E}[d(X_n, x)^2]$,，对所有 $x \in M$均成立。于是，由到任意给定点的平方距离具有 2-测地线强凸性，可得

$$\begin{aligned} V_n &\leq \mathbb{E}\left[\left(1 - \frac{1}{b}\right) d(S_{n-1}, b^*)^2 + \frac{1}{n} d(X_n, b^*)^2 - \frac{n-1}{n^2} d(S_{n-1}, X_n)^2\right] \\ &= \frac{n-1}{n} V_{n-1} + \frac{\sigma_n^2}{n} - \frac{n-1}{n^2} \mathbb{E}[F_n(S_{n-1})], \end{aligned}$$

这里我们利用了这样一个事实：根据 S_{n-1} 的构造， X_n 和 S_{n-1} 相互独立（ S_{n-1} 仅依赖于$X_1, \ldots, X_{n-1}$）。现在，再次利用到任意给定点的平方距离具有 2-测地线强凸性，可得 F_n 也具有 2-测地线强凸性，因此 $F(S_n) \geq F(b^*) + d(S_n, b^*)^2$ 几乎必然成立，因为根据假设，b^* 是 F_n 的极小值点。因此可得

$$\begin{aligned} V_n &\leq \frac{n-1}{n} V_{n-1} + \frac{\sigma_n^2}{n} - \frac{n-1}{n^2} \sigma_n^2 - \frac{n-1}{n^2} V_{n-1} \\ &= \frac{(n-1)^2}{n^2} V_{n-1} + \frac{\sigma_n^2}{n^2}. \end{aligned}$$

换言之， $n^2 V_n \leq (n-1)^2 V_{n-1} + \sigma_n^2$。由于根据 S_1 的定义， $V_1 = \sigma_1^2$，结论由归纳法得出。

□

最后，当3.8 中的 X_i不服从相同分布时，我们还有推论

定理 4.3. 设 (M, d) 为 NPC 空间，且 $X_1, \ldots, X_n$ 是 (M, d) 中的独立随机变量。假设对于某些固定的 $x_0 \in M$ 和 $C > 0$，对所有 $i = 1, \ldots, n$，均有 $d(X_i, x_0) \leq C$ 几乎必然。令 T_n 为 $X_1, \ldots, X_n$ 的 $\hat{}$ 经验重心或归纳重心。若 T_n 是 b_n 的经验重心，则进一步假设 (M, d) 的曲率有下界。那么，对所有 $\delta \in \in (0, 1)$，以下结论以至少 $1 - \delta$ 的概率成立：

$$d(T_n, b^*) \leq \frac{\overline{\sigma}_n}{\sqrt{n}} + \min\left(2\overline{\sigma}_n \sqrt{\frac{\log(1/\delta)}{n}}, \frac{8C \log(1/\delta)}{3n}\right),$$

其中 $\overline{\sigma}_n$ 与定理 *4.1*中的相同。

在定理 4.1中，即使 X_1、……、 X_n 的重心各不相同，我们仍有以下结论：对于任意固定的 $b \in M$，对于任意 $\delta \in (0, 1)$，以下结论以至少 $1 - \delta$ 的概率成立

$$d(T_n, b) \leq \mathbb{E}[d(T_n, b)] + \overline{C}_n \sqrt{\frac{\log(1/\delta)}{n}}.$$

然而，尚不清楚应选择什么 b，才能使第一项尽可能小。如果 b_1，……， b_n 分别是 X_1，……， X_n 的总体重心（即，每个 b_i 都在 $b \in M$上最小化 $\mathbb{E}[d(X_i, b)^2]$），那么 b 的一个自然候选是 b_1，……， b_n的重心。

开放问题。设 X_1，……，X_n 是 M中的独立随机变量，且具有二阶矩。对于每个 $i=1$，……，n，令 b_i 为 X_i 的重心，并令 $\sigma_i^2=\mathbb{E}[d(X_i,b_i)^2]$ 为其方差。是否有

$$\mathbb{E}[d(\hat{b}_n,b_n^*)^2]\leq\frac{\overline{\sigma}_n^2}{n},$$

其中，b_n^* 是 b_1、……、b_n 和 $\overline{\sigma}_n^2=\frac{\sigma_1^2+\ldots+\sigma_n^2}{n}$的重心，如定理4.1中所述?

5 超越非正曲率空间

很自然地会问：如果不假设非正曲率，定理 3.3 和 3.7 是否仍然成立？或许只需放宽常数和/或收敛速率。CAT 空间是研究这一问题的自然框架。简而言之，如果一个度量空间中的三角形比曲率为 κ 的模型空间中的三角形更瘦，那么称该度量空间为 CAT(κ)，$\kappa\in\mathbb{R}$（若 $\kappa<0$，则模型空间为双曲平面；若 $\kappa=0$，则为欧几里得平面；若 $\kappa>0$，则为半径为 $1/\kappa$ 的二维球面）：关于 CAT 空间的介绍，参见 [14, 15]。用这一术语来说，NPC 空间就是 CAT(0)空间。一个重要的 CAT(κ) 空间实例与最优传输理论密切相关；对于 $\kappa>0$，它是所有 $p\geq 1$阶、特征值均不小于 $\sqrt{3}/(2\kappa)$的对称正定矩阵所构成的空间，[32, 配备 Bures-Wasserstein 度量（命题 2],）。任意两个对称正定矩阵 A、B 在 $\mathbb{R}^{p\times p}$ 中的 Bures-Wasserstein 度量为 $W_2(\mathcal{N}(0,A),\mathcal{N}(0,B))$，其中 W_2 是二阶 Wasserstein 距离，而 $\mathcal{N}(0,A)$ 是协方差矩阵为 A 的 p 维中心高斯分布。有关 Bures-Wasserstein 度量的更多信息，请参阅 [10]。我们先介绍一些 CAT 空间的背景知识。

5.1 CAT 空间背景

本节将给出 CAT 空间的精确定义。关于这一主题的完整介绍，请参阅著作 [4]。首先，我们介绍一族模型空间，后文将借助它们定义局部和整体曲率界。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。该模型空间对应于零曲率，是一个测地空间，其中测地线唯一，且为线段。

$\kappa>0$：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，其中除对跖点外，测地线均唯一，且为大圆弧。这里，大圆是球面与任意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 $\mathbb{H}^2=\{(x_1,x_2,x_3)\in\mathbb{R}^3:\ x_3>0,x_1^2+x_2^2-x_3^2=-1\}$。度量由 $d_\kappa=\frac{1}{\sqrt{-\kappa}}\operatorname{arccosh}(-\kappa\langle x,y\rangle)$ 给出，对所有 $x,y\in M_\kappa$ 均成立，其中 $\langle x,y\rangle=x_1y_1+x_2y_2-x_3y_3$。这是一个测地空间，其中测地线总是唯一的，并由 M_κ 与经过原点的平面的交集弧段给出。

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

.

设 (M,d) 为测地空间，即任意两点之间至少存在一条测地线的度量空间。通过将 (M,d) 中的三角形与模型空间中边长相同的三角形进行比较，来定义曲率（下界或上界）的概念。

定义 5.1。在 M 中，一个（测地线）三角形是 M 中的三个点（顶点）以及连接它们的三条测地线段（边）所组成的集合。

给定三个点 x、y、$z\in S$，我们不严谨地将一个顶点为 x、y、z 的三角形记作 $\Delta(x,y,z)$，而不说明为其边选择了哪些测地线段（点之间的测地线不一定唯一，例如在球面上，任意两个对跖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 $\mathrm{per}(\Delta)=d(x,y)+d(y,z)+d(x,z)$。它不取决于边的选择。

定义 5.2. 设 $\kappa\in\mathbb{R}$ 和 Δ 是 M 中的一个三角形，且 $\mathrm{per}(\Delta)<2D_\kappa$。$\Delta$ 在 ¯模型空间 M_κ 中的比较三角形，是指一个与 Δ边长相同的三角形 $\Delta\subseteq M_\kappa$，即若 $\Delta=\Delta(x,y,z)$，则 $\Delta=\Delta(\bar{x},\bar{y},\bar{z})$

其中 $\bar{x}$、$\bar{y}$、$\bar{z}$ 是 M_κ 中满足以下条件的任意点

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

注意，在 M_κ 中，比较三角形在刚体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) 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（分别地，更胖），那么我们就说 (M,d) 的全局曲率上界（分别地，下界）为 κ。

定义 5.3. 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 $x, y \in \Delta$, $d(x,y) \le d_\kappa(\bar{x},\bar{y})$（分别地，$d(x,y) \le d_\kappa(\bar{x},\bar{y})$），其中 $\bar{x}$ 和 $\bar{y}$ 是模型空间 M_κ 中与 x 和y 对应的比较三角形 Δ 上的点时，我们称 (M,d) 的全局曲率上界（分别地，下界）为κ。此时称 (M,d) 为 CAT(κ)（分别地，CAT$^+$(κ)) 空间。

5.2 CAT空间中的重心

现在，下一节的目标是推导当 κ 空间满足 $\kappa > 0$ 时，经验重心的非渐近集中界。众所周知，在这类空间中，重心可能不再唯一。例如，在任意球面上，北极和南极的任意一个赤道点都是它们的重心。然而，已知如果 CAT(κ) 空间上的概率测度满足 $\kappa > 0$，并且其支撑集包含在一个足够小的球内，那么它确实具有唯一的重心 [38, 定理 B]。这与以下关于到任意固定点的平方距离的凸性性质有关 [33, 命题 3。1]。

引理 5.4。设 (M,d) 是一个 CAT(κ) 空间，满足 $\kappa > 0$，并设 $x_0 \in M$。那么，对于所有 $\varepsilon \in (0, \frac{\pi}{2\sqrt{\kappa}}]$, $\frac{1}{2}d(x_0,\cdot)^2$，其在球 $B\left(x_0, \frac{\pi}{2\sqrt{\kappa}} - \varepsilon\right)$ 上是 k_ε-测地强凸的，其中 $k_\varepsilon = (\pi - 2\sqrt{\kappa}\varepsilon)\tan(\varepsilon\sqrt{\kappa})$。

这一强　　凸性性质蕴含直径较小的 CAT(κ) 空间中关于 $\kappa >$ 的方差不等式　　0 .

引理 5.5。设 (M,d) 是一个具有 $\kappa > 0$ 的 CAT(κ) 空间，设 $x_0 \in M$，且设 $\varepsilon \in (0, \frac{\pi}{2\sqrt{\kappa}}]$。设 X 是 M 中取值于球 $B\left(x_0, \frac{\pi}{2\sqrt{\kappa}} - \varepsilon\right)$ 内的随机变量，且几乎必然成立。则它满足以下方差不等式

$$d^2(z,b^*) \le \frac{2}{k_\varepsilon}\mathbb{E}\left[d^2(z,X) - d^2(b^*,X)\right], \quad \forall z \in M, \tag{7}$$

其中 $k_\varepsilon = (\pi - 2\sqrt{\kappa}\varepsilon)\tan(\varepsilon\sqrt{\kappa})$，而 b^* 表示 X 的分布的 Fr´echet 均值。

注意，对所有 $\varepsilon \in (0, \frac{\pi}{2\sqrt{\kappa}})$, $k_\varepsilon \in (0, 2)$。此外，$k_\varepsilon \xrightarrow[\varepsilon\to 0]{} 0$，且 $k_\varepsilon \xrightarrow[\varepsilon\to \frac{\pi}{2\sqrt{\kappa}}]{} 2$。因此，当直径趋于零时，凸性趋于与欧几里得情形相同；而当直径趋于最大值（即$\frac{\pi}{2\sqrt{\kappa}}$）时，测地线强凸性可能不再成立。

证明。由 [33, 命题 3.1], 可知，对于所有 $p \in M$，函数 $z \mapsto d^2(z,p)$ 在球 $B\left(x_0, \frac{\pi}{2\sqrt{\kappa}} - \varepsilon\right)$ 上是 k_ε-凸的。此外，根据 [38, 定理 B], b^* 位于该球内。因此，函数 $z \mapsto \mathbb{E}\left[d^2(z,X) - d^2(b^*,X)\right]$ 在该球上也同样是 k_ε-凸的。因此，取 b^*与 z 之间的 z_t 连接测地线（由于球足够小，该测地线存在且唯一），并对 $p = b^*$，根据重心的定义，我们得到，对于 $t \in [0,1]$,

$$\begin{aligned} 0 &\le \mathbb{E}\left[d^2(z_t,X) - d^2(b^*,X)\right] \\ &\le t\,\mathbb{E}\left[d^2(z,X) - d^2(b^*,X)\right] - \frac{k_\varepsilon}{2}t(1-t)d^2(z,b^*) \end{aligned}$$

因此,

$$t\,\mathbb{E}\left[d^2(z,X) - d^2(b^*,X)\right] \ge \frac{k_\varepsilon}{2}t(1-t)d^2(z,b^*),$$

将两边除以 t，并令 t 趋于零，即得所需结果。 □

为了推导 $\kappa > 0$ 的 CAT(κ) 空间中经验重心收敛的非渐近界，我们需要作出以下假设。

假设 5.6. 度量空间 (M,d) 对某个 $\kappa>0$ 是 CAT(κ) 空间，并且存在正常数 $A,p>0$，使得对于所有 $x\in M$ 以及所有满足 $\alpha\le r$ 的 $\alpha,r>0$，

$$N(B(x,r),\alpha)\le\left(\frac{Ar}{\alpha}\right)^p, \tag{8}$$

其中， $N(B(x,r)\ \alpha)$表示满足以下条件的最小整数 $N\ge 1$：球 $B(x,r)$ 可以用 N个半径为 α的球覆盖。

注 5.7。 量 $N(B(x,r)\ \alpha)$ 的对数称为 $B(x,r)$ 的度量熵，并用于经验过程理论中（例如参见 [23]）。例如，如果 M 是一个维数为 p 且具有有界曲率的黎曼流形，那么它满足（8），其中相同的 p 作为指数。这可以利用黎曼流形中球的体积估计，通过体积论证轻易看出；例如参见 [16, 第三节]。因此，（8）在某种意义上刻画了 M的有限维数性。

现在可以在 CAT(κ) 空间（$\kappa>0$）的框架下陈述以下集中不等式。

定理 5.8. 设 (M,d) 为 CAT(κ) 空间， $\kappa>0$，满足假设 5.6。设 $x_0\in M$且 $\varepsilon\in(0,\frac{\pi}{2\sqrt{\kappa}}]$。令 $X_1,...,X_n$为满足 $X_1\in B\left(x_0,\frac{\pi}{2\sqrt{\kappa}}-\varepsilon\right)$几乎必然的独立同分布随机变量。记 b^*为 X_1的总体重心，并令 b_n为 $X_1,\ldots,X_n$ 的经验重心。则对于所有 $\delta\in(0,1)$，以至少 $1-\delta$的概率有

$$d(\hat{b}_n,b^*)\le c\left(\frac{A}{\varepsilon\kappa}\sqrt{\frac{p}{n}}+\frac{1}{\varepsilon^{3/2}\kappa}\sqrt{\frac{\log(2/\delta)}{n}}\right),$$

其中 $c>0$ 是一个通用常数，且 A 是假设 5.6中的常数。

证明。证明基于对[3, 定理 2.1]的应用。得益于假设 5.6 和引理 5.5，可以使用该定理。通过仔细分析其证明，并取常数 $D=p$, $C=A$,$K_1=\sqrt{\frac{\pi}{2\sqrt{\kappa}}-\varepsilon}$, $K_2=2\left(\frac{\pi}{2\sqrt{\kappa}}-\varepsilon\right)$, $K_3=\quad -{}^2(\pi-2\sqrt{\kappa}\varepsilon)\tan(\varepsilon\sqrt{\kappa})$和 $\alpha=\beta=1$，我们得到，对于所有 $\delta\in(0,1)$，以至少 $1-\delta$ 的概率，

$$\sqrt{\frac{k_\varepsilon}{2}}d(\hat{b}_n,b^*)\le 3c_1\sqrt{\frac{p}{n}}+3c_2\sqrt{\frac{\log(2/\delta)}{n}},$$

其中

$$k_\varepsilon=(\pi-2\sqrt{\kappa}\varepsilon)\tan(\varepsilon\sqrt{\kappa}),$$

$$c_1=\frac{96\sqrt{2A}\left(\frac{\pi}{2\sqrt{\kappa}}-\varepsilon\right)}{\sqrt{(\pi-2\sqrt{\kappa}\varepsilon)\tan(\varepsilon\sqrt{\kappa})}}=\frac{96\sqrt{A}\sqrt{\frac{\pi}{2\sqrt{\kappa}}-\varepsilon}}{\sqrt{2\kappa\tan(\varepsilon\sqrt{\kappa})}}$$

and

$$c_2=\frac{4\left(\frac{\pi}{2\sqrt{\kappa}}-\varepsilon\right)}{\sqrt{(\pi-2\sqrt{\kappa}\varepsilon)\tan(\varepsilon\sqrt{\kappa})}}+\frac{16}{3}\sqrt{\left(\frac{\pi}{2\sqrt{\kappa}}-\varepsilon\right)}=\sqrt{\frac{\pi-2\varepsilon\sqrt{\kappa}}{\kappa}}\left(\frac{2}{\sqrt{\tan(\varepsilon\sqrt{\kappa})}}+\frac{16}{3\sqrt{2}}\right).$$

因此，以至少 $1-\delta$ 的概率，有

$$d(\hat{b}_n,b^*)\le\frac{288\sqrt{A}}{\sqrt{\kappa}\tan(\varepsilon\sqrt{\kappa})}\sqrt{\frac{p}{n}}+\frac{1}{\sqrt{\kappa\tan(\varepsilon\sqrt{\kappa})}}\left(\frac{6\sqrt{2}}{\sqrt{\tan(\varepsilon\sqrt{\kappa})}}+16\right)\sqrt{\frac{\log(2/\delta)}{n}},$$

注意到对于所有 $x\in[0,\pi/2)$，$\tan x\ge x$，即可证明该结论。 □

致谢：第二作者谨向 Mikl´os P´alfia 和 Takomi Yokota 致谢，感谢他们提供的有益讨论。

参考文献

[1] Bijan Afsari. 黎曼 l^p-质心：存在性、唯一性与凸性。《美国数学会会刊》，139(2):655–673，2011。

[2] Martial Agueh 和 Guillaume Carlier。《Wasserstein 空间中的重心》。《SIAM 数学分析杂志》，43(2):904–924，2011年。

[3] A. Ahidar-Coutrix、T. Le Gouic 和 Q. 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年。[4]Stephanie Alexander、Vitali Kapovitch 和 Anton Petrunin。《亚历山德罗夫几何导论》。《数学简明教程丛书》。施普林格，瑞士查姆，2019年。CAT(0)空间。[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年。[6] Jason M Altschuler 和 Enric Boix-Adsera。《计算 Wasserstein重心是 NP困难的》。《SIAM 数据科学数学杂志》，4(1):179–203，2022年。[7] Miroslav Bac´ak。《Hadamard空间中的凸分析与优化》。载于《Hadamard空间中的凸分析与优化》。德·格鲁伊特出版社，2014年。[8] Rajendra Bhatia。《正定矩阵》。载于《正定矩阵》。普林斯顿大学出版社，2009年。[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年。[10] Rajendra Bhatia、Tanvi Jain 和 Yongdo Lim。《正定矩阵之间的 Bures–Wasserstein距离》。《数学论述》，37(2):165–191，2019年。[11] Rabi Bhattacharya 和 Lizhen Lin。《Fr´échet 均值的综合中心极限定理与非欧空间上的非参数推断》。《美国数学会会刊》，145(1):413–428，2017年。[12] Rabi Bhattacharya 和 Vic Patrangenaru。《流形上内在与外在样本均值的大样本理论》。《统计年刊》，31(1):1–29，2003年。[13] Rabi Bhattacharya 和 Vic Patrangenaru。《流形上内在与外在样本均值的大样本理论：II》。《统计年刊》，第1225–1259页，2005年。[14] Martin R Bridson 和 Andr´e Haefliger。《非正曲率度量空间》，第319卷。施普林格科学与商业媒体，2013年。[15] Dmitri Burago、Yuri Burago 和 Sergei Ivanov。《度量几何教程》，第33卷。美国数学学会，2022年。[16] Isaac Chavel。《黎曼几何：现代导论》，第98卷。剑桥大学出版社，2006年。[17] Sebastian Claici、Edward Chien 和 Justin Solomon。《随机Wasserstein重心》。载于《国际机器学习大会》，第999–1008页。机器学习研究论文集，2018年。[18] Marco Cuturi 和 Arnaud Doucet。《快速计算 Wasserstein重心》。载于《国际机器学习大会》，第685–693页。机器学习研究论文集，2014年。[19] Benjamin Eltzner、Fernando Galaz-Garcia、Septhan F Huckemann 和 Wilderich Tuschmann。《割点集的稳定性及黎曼流形 Fr´échet 均值的中心极限定理》。arXiv预印本 arXiv:1909.00410，2019年。[20] Benjamin Eltzner 和 Stephan F Huckemann。《流形上的涂抹型中心极限定理及其在高维球面上的应用》。《统计年刊》，47(6):3360–3381，2019年。[21] Maurice Fr´échet。《距离空间中任意性质的随机元》。《亨利·庞加莱研究所年刊》，10:215–310，1948年。[22] Kei Funano。《取值于 $\mathbb{R}$-树和 Hadamard流形的随机过程的收敛速率》。《大阪数学杂志》，47(4):911–920，2010年。[23] Evarist Gin´é。《经验过程及其应用：综述》。《伯努利》，2(1):1–38，1996年。附有 Jon A. Wellner 的讨论及作者的回应。[24] Thomas Hotz、Sean Skwerer、Stephan Huckemann、Huiling Le、J Stephen Marron、Jonathan C Mattingly、Ezra Miller、James Nolen、Megan Owen 和 Vic Patrangenaru。《开书空间上的黏性中心极限定理》。《应用概率年刊》，2013年。[25] 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年。

[26] 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年。[27] Alexey Kroshnin、Nazarii Tupitsa、Darina Dvinskikh、Pavel Dvurechensky、Alexander Gasnikov 和 Cesar Uribe。《关于近似 Wasserstein 重心的复杂度》。载于《国际机器学习大会》，第3530–3540页。机器学习研究论文集，2019年。[28] T. Le Gouic、Q. Paris、P. Rigollet 和 A.J. Stromme。《亚历山德罗夫空间和 Wasserstein 空间中经验重心的快速收敛》。欧洲数学会杂志，2022年。[29] 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年。[30] Michel Ledoux。《测度集中现象》，数学调查与专著第89卷。美国数学学会，罗得岛州普罗维登斯，2001年。[31] Yongdo Lim 和 Mikl´os P´alfia。《Hadamard 空间中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。[32] 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《几何信息科学：第四届国际会议 GSI 2019，法国图卢兹，2019年8月27日至29日，会议论文集》，第739–748页。施普林格，2019年。[33] Shin-ichi Ohta。《度量空间的凸性》。几何奉献，125:225–250，2007年。[34] 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》。变分法与偏微分方程，54(2):1591–1610，2015年。[35] Barry Simon。《单调矩阵函数的 Loewner 定理》。施普林格，2019年。[36] 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年），当代数学第338卷，第357–390页。美国数学学会，罗得岛州普罗维登斯，2003年。[37] Roman Vershynin。《高维概率：简介及其在数据科学中的应用》，第47卷。剑桥大学出版社，2018年。[38] Takumi Yokota。《小半径 CAT(1)空间上的凸函数与重心》。日本数学会杂志，68(3):1297–1323，2016年。[39] Takumi Yokota。《小半径 CAT(1)空间上的凸函数与 p-重心》。筑波数学杂志，41(1):43–80，2017年。[40] 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。机器学习研究论文集，2016年。[41]Hongyi Zhang 和 Suvrit Sra。《迈向黎曼加速梯度方法》。arXiv预印本 arXiv:1806.02812，2018年。[42] Herbert Ziezold。《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。施普林格，1977年。