宽型随机神经网络深度极限下的大偏差原理和函数型极限定理*

Simmaco Di Lillo† Claudio Macci‡ Barbara Pacchiarotti§

2026年1月9日

arXiv:2601.04677v1 [math.PR] 8 Jan 2026

摘要

本文研究了一类定义在单位球面上的连续、各向同性、中心化高斯随机场的大偏差原理和弱收敛，包括有限维分布层面和函数形式。这些场的协方差函数通过激活函数诱导的非线性映射递归演化，反映了随着深度增加，宽度趋于无穷的随机神经网络的统计动力学。我们考虑两类中心化场，它们分别通过减去北极处的值或球面平均得到。根据相关协方差函数在 $t = 1$ 处导数的行为，我们识别出三种状态：低无序、稀疏和高无序。在低无序区间，我们建立了函数型大偏差原理和弱收敛结果。在稀疏区间，我们得到有限维分布的大偏差原理和弱收敛；但由于协方差递归中出现不连续性，这两种性质在函数层面均不成立。

关键词 **词：** 高斯随机场、各向同性、稀有事件渐近、谱复杂度 ，
球面 几何、过程的弱收敛。
*2010*年数学主题分类：60F10、60G15、60G60、68T07。

1 引言

近年来，对神经网络的概率分析揭示了当每层神经元数量趋于无穷时，它们的渐近行为。在这一极限下，深层前馈随机网络的输出收敛于高斯过程，其协方差函数按照由激活函数和初始化参数决定的非线性递归关系演化。在这一机制中（通常称为核区域或无限宽极限），前馈神经网络可用高斯过程进行概率表示，而这些过程完全由其对应的正定核刻画 [5,9,12,17,18,21,24,27–29,32]。此外，与深层架构相对应的核具有递归复合结构：它可以写成固定基核的 L次迭代，其中 L表示网络深度。高斯随机场与深层神经网络架构之间的联系，促使多项研究致力于理解高维空间中神经表示的统计与几何性质。

*作者感谢 MUR 卓越系部项目（授予罗马第二大学数学系，CUP E83C23000330006）、罗马第二大学 METRO 研究项目（CUP E83C25000630005）以及 INdAM-GNAMPA 的支持。

†地址：罗马第二大学数学系，科学研究路，I-00133 意大利罗马。电子邮件： dilillo@mat.uniroma2.it

‡地址：罗马第二大学数学系，科学研究路，I-00133 意大利罗马。电子邮件： macci@mat.uniroma2.it§地址：罗马第二大学数学系，科学研究路，I-00133 意大利罗马。电子邮件： pacchiar@mat.uniroma2.it

在这一背景下，一个自然的问题是理解这些随机场随深度增加而产生的渐近涨落。针对这一问题，近期研究考察了平均场机制中相互作用神经元模型和随机神经网络的大偏差原理，为经验过程或递归高斯系统建立了大偏差原理，并提供了严谨的概率框架，以量化高维神经模型中的罕见涨落（例如，参见 [2, 19, 25, 31]）。这些进展促使我们开展本研究。

众所周知（例如，见 [16]），深度随机神经网络的定性行为由迭代协方差函数的性质决定。特别是，根据导数 $\kappa'(1)$ 的取值，可以区分出称为低无序、稀疏和高无序 的几个区间。我们的分析主要受到 [16], 中定理 3.3 的启发，该定理证明，在低无序和稀疏区间中，以北极为中心化点后，相应高斯场收敛于零。在本文中，我们旨在将这些结果纳入大偏差框架加以深化，并进一步考察另一种中心化方式，即减去场的球面平均。

本文针对所考虑的两种中心化方式，分别在低无序和稀疏区间内建立有限维分布的大偏差原理。我们还证明了一些有限维弱收敛结果。这两个区间之间存在一个根本区别：在低无序情形下，我们得到函数型大偏差原理和函数型弱收敛；而在稀疏情形下，由于协方差函数出现不连续性，这两者在函数层面上均不成立。我们还研究了高无序区间。低无序与稀疏情形之间的差异，可归因于随着深度增加，网络初始化所占的相对权重不同。在低无序区间，高斯场会以指数速度趋于平凡，这既体现在有限维分布层面，也体现在函数空间中。相比之下，在稀疏区间，衰减速度为多项式级；尽管逐点收敛于零仍然成立，但初始化的权重使函数型收敛结果无法成立。这种差异反映出更深层的结构性对比：在低无序区间，深度会一致地抹去信息；而在稀疏情形下，其影响仍然可见。

本文结构如下。第 2 节回顾主要的概率论与分析预备知识。第 3 节和第 4 节分别介绍有限维结果和函数型结果。第 5 节讨论高无序区间。为便于阅读，所有证明均放在附录 A 中。

总的来说，本文将高斯随机场的经典渐近理论拓展至递归与非线性情形，并阐明支配深度如何影响宽型随机神经网络行为的概率机制。通过将谱观测与概率极限联系起来，我们提供了一个框架，用于量化信号在极深层架构中的传播退化，并理解深度和初始化区间如何影响随机神经网络的表达与统计性质。该框架也为未来涉及高阶场的研究奠定了基础，例如纳入导数[16]的研究，以及向非高斯模型的拓展。

2 预备知识

本节介绍大偏差、球面上的各向同性随机场以及随机神经网络的一些预备知识。

2.1 大偏差

我们将 [13] 作为这一主题的参考文献。我们先介绍大偏差原理（下文简称 LDP）的定义。设一列随机变量$\{X_L\}^{L\geq 1}$ （取值于某个拓扑空间$\mathcal{X}$，并定义在同一概率空间 $(\Omega, \mathcal{F}\ P)$ 上），令 $L \to \infty$,

因为这种表述更直接，也更适用于下文。下半连续函数 $I:\mathcal{X} \to [0,\infty]$ 称为速率函数；若其所有水平集$\{\{x\in\mathcal{X}:\ I(x)\le\eta\}:\ \eta\ge 0\}$ 都是紧致的，则称其为良好的。若$\{X_L\}_{L\ge 1}$ 以速度 $v_L\to\infty$和速率函数 I 满足大偏差原理，则

$$\limsup_{L\to\infty}\frac{1}{v_L}\log P(X_L\in C)\le -\inf_{x\in C} I(x) \text{ for all closed sets } C\subset\mathcal{X}$$

and

$$\liminf_{L\to\infty}\frac{1}{v_L}\log P(X_L\in O)\ge -\inf_{x\in O} I(x) \text{ for all open sets } O\subset\mathcal{X}.$$

一般而言，我们研究中心高斯序列。我们将证明有限维分布的大偏差结果，并考虑 G¨artner–Ellis 定理的一些应用（例如参见 [13] 中的定理 2.3.6）。我们还将给出定义在球面上的连续随机场序列的函数型结果。这里，速率函数将表示为有限维分布速率函数的适当上确界（例如参见 [6]）；从定理 4.1 的陈述中可以清楚地看出这一点。最后，我们还将引用收缩原理（例如参见 [13] 中的定理 4.2.1）以及 Dawson G¨artner 定理（ [13] 中的定理 4.6.1）。

2.2 球面上的各向同性随机场

在本文中，我们使用符号 $C(\mathbb{S}^d)$ 表示 $\mathbb{S}^d$ 上所有取实值的连续函数所组成的空间。我们回顾定义在概率空间 $(\Omega,\ \mathcal{F}\ P)$ 上的连续、各向同性中心高斯随机场 $(T(x))_{x\in\mathbb{S}^d}$ 的一些众所周知的性质；例如参见 $[26,33]$。我们用 λ 表示 $\mathbb{S}^d$ 上的勒贝格测度，用 $\omega_d=\lambda(\mathbb{S}^d)$ 表示其总质量，并采用以下谱表示

$$T(x,\omega)=a_{00}(\omega)Y_{00}(x)+\sum_{\ell=1}^{\infty}\sum_{m=1}^{n_{\ell,d}} a_{\ell m}(\omega)\,Y_{\ell m}(x), \tag{1}$$

这在 $L^2(\Omega\times\mathbb{S}^d)$ 中成立，即

$$\lim_{\ell_*\to\infty}\int_\Omega\int_{\mathbb{S}^d}\left|T(x,\omega)-\left(a_{00}(\omega)Y_{00}(x)+\sum_{\ell=1}^{\ell_*}\sum_{m=1}^{n_{\ell,d}} a_{\ell m}(\omega)Y_{\ell m}(x)\right)\right|^2\lambda(dx)\,P(d\omega)=0.$$

这里：

- $n_{\ell,\ d}=\underline{\qquad\frac{2\ell+d-1}{\ell}\qquad}\binom{\ell+d-2}{\ell-1}$ 当 $\ell\ge 1$ 时；
- $\mathcal{A}=\{(0,\ 0)\}\cup\{(\ell,\ \ m):\ \ \ell\ge 1,\ \ \ m=1,\ \ldots\ldots,\ \ n_{\ell,\ d}\}$；
- $\{a_{\ell m}\}_{(\ell,\ m)\in\mathcal{A}}$ 是相互独立的中心高斯随机变量，且满足 $\mathbb{E}[a_{\ell m}^2]=C_\ell\ge 0$；
- $\{Y_{\ell m}\}_{(\ell,\ m)\in\mathcal{A}}$ 是一个 $L^2(\lambda)$ 实值球谐函数的正交归一基，其中每个 $Y_{\ell m}$都是拉普拉斯-贝尔特拉米算子的特征函数，对应的特征值为 $-\ell(\ell+d-1)$。

由谱表示和正交归一性可立即得到

$$\frac{1}{\omega_d}\int_{\mathbb{S}^d}T(x)\,\lambda(dx)=\frac{a_{00}}{\sqrt{\omega_d}},$$

因为 $Y_{00}(x)\ =Y_{00}\equiv\omega_d^{-1/2}$。

此外，各向同性意味着存在一个连续函数 $\kappa:\ [-1,1]\to\mathbb{R}$ 使得

$$\mathrm{Cov}(T(x),T(y))=\kappa(\langle x,y\rangle)\qquad \text{for all } x,y\in\mathbb{S}^d, \tag{2}$$

特别地，$\mathsf{Var}[T(x)]=\kappa(1)$。利用球谐函数的加法公式，[3], 协方差函数可展开为

$$\kappa(t)=\sum_{\ell\geq 0}\frac{C_\ell n_{\ell,d}}{\omega_d}\,G_{\ell,d}(t),$$

其中 $G_{\ell,\ d}$ 是次数为 ℓ 的归一化盖根鲍尔多项式，由取值 $G_{\ell,\ d}(1)=1$ 以及正交关系唯一确定（见 [30]）

$$\int_{-1}^{1}G_{\ell,d}(t)\,G_{\ell',d}(t)\,(1-t^2)^{\frac{d}{2}-1}\,dt=0,\qquad \ell\neq\ell'.$$

为方便起见，记

$$D_\ell=\frac{C_\ell n_{\ell,d}}{\omega_d},$$

从而协方差满足

$$\kappa(1)=\sum_{\ell\geq 0}D_\ell.$$

2.3 随机神经网络

我们回顾一些关于随机神经网络的预备概念；例如，参见 [16] 的第 2.2 节和命题 A.6。随机神经网络递归定义如下。对于输入层，设

$$Z_{0,i}(x)=\sum_{j=1}^{m_0}W_{ij}^{(0)}x_j+b_i^{(1)},\qquad i=1,\dots,m_1,$$

且，对于 $s=1,\dots,L$,

$$Z_{s,i}(x)=\sum_{j=1}^{m_s}W_{ij}^{(s)}\,\sigma(Z_{s-1,j}(x))+b_i^{(s+1)},\qquad i=1,\dots,m_{s+1}.$$

上面出现的随机变量相互独立，且服从适当的中心化高斯分布：偏置 $b_i^{(s)}$，$s=1,\dots,L$, 的方差为 $\Gamma_b\geq 0$，初始权重 $W_{ij}^{(0)}$ 的方差为$\Gamma_{W_0}>0$，而权重 $W_{ij}^{(s)}$（$s\geq 1$）的方差为 $\Gamma_W/m_s>0$。我们用 σ 表示：$\mathbb{R}\to\mathbb{R}$为激活函数；L 为隐藏层总数；$m_0=d+1$ 为输入维度，m_{L+1}为输出维度，而 $m_1,\dots,m_L$ 为各隐藏层的宽度。

随机神经网络的宽度极限，即当 $m_1,\dots,m_L\to\infty$ 时的极限，已在许多文献中得到研究；例如参见 [5, 9, 17, 18, 28] 及其中的参考文献。更确切地说，可以证明这一族

$$\{(Z_{L,i}(x))_{x\in\mathbb{S}^d}\}_{i=1,\dots,m_{L+1}}$$

随着宽度趋于无穷，依分布收敛于分量独立同分布的高斯随机向量

$$\{(Z_{L,i}^*(x))_{x\in\mathbb{S}^d}\},$$

其协方差函数可以关于深度 L递归定义。特别地，由于极限分量是独立同分布的，我们只需研究一个单变量极限分量，将其记为（$T_L(x)$)$_{x\in\mathbb{S}^d}$。

在标准校准条件下，

$$\mathbb{E}[\sigma(Z)^2] < \infty, \qquad \Gamma_b + \Gamma_{W_0} = 1, \qquad \Gamma_W \, \mathbb{E}[\sigma(Z)^2] + \Gamma_b = 1,$$

对于标准正态随机变量 Z，极限协方差函数为

$$\mathrm{Cov}(T_L(x), T_L(y)) = \kappa_L(\Gamma_b + \Gamma_{W_0} \langle x, y \rangle),$$

其中 κ_L 表示 κ 的 L 重复复合，即

$$\kappa_L = \underbrace{\kappa \circ \cdots \circ \kappa}_{L \text{ times}}$$

and

$$\kappa(t) = \Gamma_W \, \mathbb{E}\Big[\sigma(\widehat{Z}_1) \, \sigma\big(t \, \widehat{Z}_1 + \sqrt{1-t^2} \, \widehat{Z}_2\big)\Big] + \Gamma_b, \qquad t \in [-1, 1]$$

$\widehat{\ }$ $\widehat{\ }$，其中 Z_1, Z_2 为独立同分布的标准正态随机变量。更多详情请参见 [16] 。若 σ 属于 Malliavin–Sobolev空间，则它可表示为 Hermite展开，即

$$\sigma(x) = \sum_{q \geq 0} \frac{\widehat{J}_q(\sigma)}{q!} \, H_q(x), \qquad \sum_{q \geq 1} \frac{|\widehat{J}_q(\sigma)|}{(q-1)!} < \infty.$$

利用图公式（参见 [26, 命题 4.15]），可得

$$\kappa(t) = \sum_{q \geq 0} \frac{(J_q(\sigma))^2}{q!} \, t^q, \quad J_q(\sigma) = \begin{cases} \sqrt{\Gamma_W \widehat{J}_0(\sigma)^2 + \Gamma_b} & \text{if } q = 0, \\ \sqrt{\Gamma_W} \widehat{J_q}(\sigma) & \text{otherwise.} \end{cases} \tag{3}$$

我们进一步假设 σ 不是多项式；否则，每一层都会对输入进行多项式变换，而这些变换的复合仍为多项式，从而无法实现通用近似（参见 [11, 20]）。等价地，对于无穷多个 q 值， $J_q(\sigma) \neq 0$ 。上述可和性条件保证了 $\kappa \in C^1([-1, 1])$。

由 n 此后，为清晰起见，我们仅限于情形 $\Gamma_b = 0$。然后 $\Gamma_{W_0} = 1$ 和

$$\Gamma_W = \mathbb{E}[\sigma(Z)^2]^{-1}$$

尽管本文结果的某个版本也可针对一般情形 $\Gamma_b \in (0,1)$ 加以建立，但相应的记号与论证会变得复杂得多。

2.4 谱复杂度的三种状态

受作为随机神经网络的无限宽极限而出现的高斯场结构的启发，本文将聚焦于一类连续、各向同性的中心高斯随机场数列 $\{(T_L(x))_{x \in \mathbb{S}^d}\}_{L \geq 1}$, 其中，对每一个 $L \geq 1$,

$$\mathrm{Cov}(T_L(x), T_L(y)) = \kappa_L(\langle x, y \rangle), \qquad x, y \in \mathbb{S}^d. \tag{4}$$

这里 $\kappa : [-1, 1] \to \mathbb{R}$ 是一个 $C^1([-1, 1])$ 函数，它具有 (3) 中的幂级数表示，且对无穷多个 q 有 $J_q(\sigma) \neq 0$ ，并经规范化使得

$$\kappa(1) = 1.$$

此外， κ_L 仍表示 κ 的 L 重复合。

研究了场 $\{(T_L(x))_{x \in \mathbb{S}^d}\}_{L \geq 1}$ 在 $L \to \infty$ 时的渐近行为 [16], ，其中根据协方差核的导数在 1 处的值，仅确定了三种状态：

- 低无序，若 $\kappa'(1) < 1$;
- 稀疏，若 $\kappa'(1) = 1$;
- 高无序，若 $\kappa'(1) > 1$。

注 2.1。 量 $\kappa'(1)$ 决定了邻近点处协方差函数的局部行为，因而控制相关性的衰减速率。它也与梯度方差直接相关；特别是，它决定了场的局部光滑性。此外，在 ***Kac–Rice*** 型公式中， $\kappa'(1)$ 在临界点或零点集期望数量的渐近分析中起着关键作用。因此， $\kappa'(1)$ 充当了随机场的谱性质与其几何或拓扑特征之间的桥梁。 [14, 15]中对这些方面进行了详细讨论；另请参阅经典文献 [1, 4]。

我们的目标是从大偏差的角度研究文献中已确立的某些趋于零的收敛现象（见引理 2.1）。更确切地说，我们将引理 2.1 中出现的数列视为随机场，即令 x 遍历 $\mathbb{S}^d$。我们首先考虑任意有限点集 $x_1, \ldots\ldots, x_m \in \mathbb{S}^d$，其中 $x_i \neq x_j$对 $i \neq j$成立，且 $m \geq 1$。在这一设定下，我们研究取值于 $\mathbb{R}^m$ 并收敛到零向量 0 的数列；或者从函数观点来看，研究在 $\mathbb{S}^d$上收敛到恒零函数的场数列。

引理 2.1. 若 $\kappa'(1) \leq 1$，则对每个固定的 $x \in \mathbb{S}^d$，当 $L \to \infty$ 时，以下数列在 $L^2(\Omega)$ 中收敛于零:

1. $\{T_L(x) - T_L(N)\}_{L\geq 1}$，其中 N 表示 $\mathbb{S}^d$ 的北极；

2. $$\left\{T_L(x) - \frac{1}{\omega_d}\int_{\mathbb{S}^d} T_L(y)\,\lambda(dy)\right\}_{L\geq 1} = \left\{T_L(x) - \frac{a_{00,L}}{\sqrt{\omega_d}}\right\}_{L\geq 1}.$$

证明. 第 1 项中的收敛性由定理 3.3 和 [16] 中的恒等式 (3.7) 得出。第 2 项中的收敛性则是 [16] 表 1 所总结结果的推论。 □

3 有限维分布的结果

本节考虑引理 2.1 中引入的数列，并研究它们的有限维分布。

在全文中，我们令

$$\underline{x} = (x_1, \ldots, x_m) \in (\mathbb{S}^d)^m,$$

将球面上两两不同点的配置记为前文所述的形式。我们还记

$$T_L(\underline{x}) = \big(T_L(x_1), \ldots, T_L(x_m)\big) \in \mathbb{R}^m,$$

并用 $\mathbf{1} \in \mathbb{R}^m$ 表示坐标均为 1 的向量。

本文的一项基本假设涉及协方差函数 κ 的正则性。在陈述主要结果之前，我们先明确表述这一条件。

假设 A (协方差正则性)。 协方差函数 $\kappa : [-1, 1] \to \mathbb{R}$ 满足，当$t \to 1^-$时，

$$1 - \kappa(t) = \kappa'(1)(1 - t) - c(1 - t)^\rho + o\big((1 - t)^\rho\big), \tag{5}$$

对于某个 $c \in \mathbb{R} \setminus \{0\}$ 以及某个 $\rho > 1$。

注3.1. 若干文献中都出现了这类假设；例如，[8]（定理 *1* 以及第 *3.2* 节中 *ReLU*网络的显式公式），以及 [15]（定义 *3.1*，其中指数 ρ 称为协方差正则性指数）。后一篇文献还进一步证明，只要 $\rho > 1$，相应的极限场几乎必然属于 $C^1(\mathbb{S}^d)$类。

我们强调，在低无序区间，关于有限维分布的大偏差原理不需要假设 *A*。不过在这种情况下，我们无法排除相应的速率函数可能是平凡的（详情见注 *3.2*）。相比之下，在稀疏区间，假设 *A* 至关重要。

现在我们可以陈述大偏差原理了。下面的两个定理各自确立两个不同的大偏差原理，分别对应稀疏区间和低无序区间。在这两个定理中，都会出现两个极限函数，每个区间对应一个。引理 3.2 和 3.3 分别针对低无序区间和稀疏区间，并保证了这些极限函数的存在。

定理 3.1。设 $\kappa'(1) \leq 1$；此外，如果 $\kappa'(1) = 1$，我们还假设假设 *A*。则数列

$$\left\{ T_L(\underline{x}) - T_L(N)\,\mathbf{1} \right\}_{L\geq 1}$$

满足速度为 v_L、良好速率函数为 $I_{1;\mathfrak{g}}(\cdot;x)$ 的大偏差原理，其中速率函数由下式给出：

$$I_{1;\mathfrak{g}}(\underline{y};\underline{x}) = \sup_{\underline{\theta}\in\mathbb{R}^m} \left\{ \langle \underline{\theta}, \underline{y}\rangle - \frac{1}{2}\langle \underline{\theta}, B_{1;\mathfrak{g}}(\underline{x})\,\underline{\theta}\rangle \right\}, \quad \underline{y}\in\mathbb{R}^m, \tag{6}$$

其中 $\mathfrak{g}$： $[-1,1]\to\mathbb{R}$, $B_{1;\mathfrak{g}}(x)$ 是矩阵

$$B_{1;\mathfrak{g}}(\underline{x}) = \left(\mathfrak{g}(\langle x_i, N\rangle) + \mathfrak{g}(\langle x_j, N\rangle) - \mathfrak{g}(\langle x_i, x_j\rangle)\right)_{i,j=1}^m, \tag{7}$$

and:

- *if* $\kappa'(1) < 1$, *then* $v_L = (\kappa'(1))^{-L}$ *and* $\mathfrak{g} = \mathfrak{L}$, *where*

$$\mathfrak{L}(t) := \lim_{L\to\infty} (\kappa'(1))^{-L}(1-\kappa_L(t));$$

- 若 $\kappa(1) = 1$，则 $v_L = L^{1/(\rho-1)}$ 和 $\mathfrak{g} = \mathfrak{S}$，其中

$$\mathfrak{S}(t) := \lim_{L\to\infty} L^{1/(\rho-1)}(1-\kappa_L(t)).$$

定理 3.2. 在与定理 *3.1* 相同的假设下，序列 e

$$\left\{ T_L(\underline{x}) - \frac{a_{00,L}}{\sqrt{\omega_d}}\mathbf{1} \right\}_{L\geq 1}$$

满足速度为 v_L 的大偏差原理，其良好速率函数为 $I_{2;\mathfrak{g}}(\cdot;x)$，由下式给出

$$I_{2;\mathfrak{g}}(\underline{y};\underline{x}) = \sup_{\underline{\theta}\in\mathbb{R}^m} \left\{ \langle \underline{\theta}, \underline{y}\rangle - \frac{1}{2}\langle \underline{\theta}, B_{2;\mathfrak{g}}(\underline{x})\underline{\theta}\rangle \right\}, \quad \underline{y}\in\mathbb{R}^m, \tag{8}$$

其中 $B_{2;\mathfrak{g}}(\underline{x})$ 是矩阵

$$B_{2;\mathfrak{g}}(\underline{x}) = \left(\frac{\int_{-1}^1 \mathfrak{g}(t)\,(1-t^2)^{\frac{d}{2}-1}\,dt}{\int_{-1}^1 (1-t^2)^{\frac{d}{2}-1}\,dt} - \mathfrak{g}(\langle x_i, x_j\rangle) \right)_{i,j=1}^m, \tag{9}$$

并且 v_L 和 $\mathfrak{g}$ 与定理 *3.1* 中的一致。

引理 3.2 和 3.3 表明，函数 $\mathfrak{L}$ 和 $\mathfrak{S}$ 定义良好。我们还证明，极限函数仅在被 κ映射到 1 的点集上取零，即

$$\mathcal{I}(\kappa) := \{t \in [-1,1] : \kappa(t) = 1\}.$$

下面的引理对该集合进行了完整刻画。

引理3.1。 假设 $\kappa'(1) \le 1$。那么只可能出现两种情况：要么 $I(\kappa) = \{1\}$，要么 $I(\kappa) = \{-1,1\}$。此外，当且仅当 κ 是对称的，即对所有 $t \in [-1,1]$，都有 $\kappa(t) = \kappa(-t)$ 时，$I(\kappa) = \{-1,1\}$。

现在我们可以陈述这两个极限函数的存在性。

引理 3.2。 假设 $\kappa'(1) < 1$。那么极限

$$\mathfrak{L}(t) = \lim_{L\to\infty} (\kappa'(1))^{-L}(1 - \kappa_L(t))$$

在 $[0,\infty)$ 中存在。此外，在假设 A 下，当且仅当 $t \in I(\kappa)$ 时，$\mathfrak{L}(t) = 0$。

引理 3.3。 假设 $\kappa'(1) = 1$ 且假设 A 成立。那么

$$\mathfrak{S}(t) = \lim_{L\to\infty} L^{1/(\rho-1)}(1 - \kappa_L(t)) = \begin{cases} 0, & t \in \mathcal{I}(\kappa), \\ h_{c,\rho}, & t \notin \mathcal{I}(\kappa), \end{cases}$$

其中 $h_{c,\rho} = (c(\rho-1))^{-1/(\rho-1)}$，且 $c > 0$ 是 $(1-t)^{\rho}$ 在 (5) 中的系数。

注 3.2 （假设 A 避免了低无序情形中的平凡性）。 假设 $\kappa'(1) < 1$。我们注意到，速率函数 $I_{1,\mathfrak{L}}$ 和 $I_{2,\mathfrak{L}}$ 在 $y = 0 \in \mathbb{R}^m$ 处均为零，因此无法先验排除所得速率函数 I 是平凡的可能性，即

$$I(\underline{y}) = \begin{cases} 0, & \underline{y} = \underline{0}, \\ \infty, & otherwise. \end{cases}$$

如果矩阵 $B_{1,\mathfrak{L}}$ 和$B_{2,\mathfrak{L}}$ 具有非零秩，则不会出现这种情形。现在我们证明，在假设 (5) 下，对于 $m > 2$，这两个矩阵必然都包含非零项。

1. 矩阵 $B_{1,\mathfrak{L}}$的对角项为 $2\mathfrak{L}(\langle x_i, N\rangle)$，其中 $i = 1,\dots,m$。若 $x_i \notin \{S,N\}$，其中 S表示南极，则 $\langle x_i, N\rangle \neq \pm 1$，根据引理 *3.2* 的第二部分，我们有$\mathfrak{L}(\langle x_i, N\rangle) > 0$。因此，$B_{1,\mathfrak{L}}$中相应的对角项非零。

2. 矩阵 $B_{2,\mathfrak{L}}$ 的对角项等于

$$\frac{\int_{-1}^{1} \mathfrak{L}(t)(1-t^2)^{\frac{d}{2}-1}\,dt}{\int_{-1}^{1}(1-t^2)^{\frac{d}{2}-1}\,dt}.$$

由于对于 $(-1,1)$ 中所有的 t 和 $\mathfrak{L} \neq 0$，$\mathfrak{L}(t) \ge 0$，该量严格为正，因此每个对角项都非零。

由于在稀疏情形下，极限函数 $\mathfrak{S}$具有特别简单的形式，因此可以推导出两个速率函数 $I_{1;\mathfrak{S}}$ 和 $I_{2;\mathfrak{S}}$ 的显式表达式。这些公式对于说明在稀疏区间内，上述两个定理不具有函数对应形式也将发挥重要作用（见定理4.2）。

注3.3 （ $I_{1;\mathfrak{S}}$的显式表述）。 我们区分两种情形。

- 首先，假设 κ 不是对称的，因此 $I(\kappa) = \{1\}$。若点 x_1、……、 x_m中没有一个与北极 N重合，则矩阵 $B_{1;\mathfrak{S}}(x) = (b_{ij}(x))_{i,j=1}^{m}$ 为

$$b_{ij}(\underline{x}) = \begin{cases} 2h_{c,\rho}, & i = j, \\ h_{c,\rho}, & i \neq j, \end{cases}$$

因此可逆。可以验证，对于 $y = (y_1, \ldots, y_m)$，有

$$I_{1;\mathfrak{S}}(\underline{y}; \underline{x}) = \frac{1}{2(m+1)h_{c,\rho}} \left(m \sum_{i=1}^{m} y_i^2 - 2 \sum_{1 \leq i < j \leq m} y_i y_j \right). \tag{10}$$

若 $x_i = N$对某个 $i \in \{1, \ldots, m\}$ 成立（我们假设 $m \geq 2$ 以避免平凡情形），令 y_i和x_i 分别表示从 y 和 x中去掉第 i 个分量后得到的向量。在这种情况下，问题可化简为 $m-1$ 个点的情形，并且

$$I_{1;\mathfrak{S}}(\underline{y}; \underline{x}) = \begin{cases} I_{1;\mathfrak{S}}(\underline{y}_i; \underline{x}_i), & y_i = 0, \\ \infty, & otherwise. \end{cases}$$

- 现在假设 κ 是对称的，因此 $I(\kappa) = \{-1, 1\}$。只要 x 中任意两点不互为对跖点，公式 (10) 仍然成立。若 $x_i = N$ 或 $x_i = S$，且其对跖点不在 x_1、……、 x_m之中，则同样可以像上文一样将问题化简为 $m-1$ 个点的情形。

另一方面，若存在 $2v$ 对径点，例如

$$x_{i_1} = -x_{j_1}, \ \ldots, \ x_{i_v} = -x_{j_v},$$

令 y_{i_v}和 x_{i_v}表示去掉指标为 $i_1, \ldots, i_v$的分量后得到的向量。于是问题化简为一个含有 $m - v$ 个点的系统，并且

$$I_{1;\mathfrak{S}}(\underline{y}; \underline{x}) = \begin{cases} I_{1;\mathfrak{S}}(\underline{y}_{i_v}; \underline{x}_{i_v}), & y_{i_1} = y_{j_1}, \ldots, y_{i_v} = y_{j_v}, \\ \infty, & otherwise. \end{cases}$$

最后，若对于某个 $h \in \{1, \ldots, v\}$，有 $(x_{i_h}, x_{j_h}) = (N, S)$ 或 (S, N)，那么问题可化简为 $m - v - 1$个点：必须施加 $y_{i_h} = y_{j_h} = 0$；否则， $I_{1;\mathfrak{S}}(y_{i_v}; x_{i_v}) = \infty$。

注3.4 (对 $I_{2;\mathfrak{S}}$ 的显式表述)。 出现在 (9)中的矩阵 $B_{2,\mathfrak{S}}(x) = (b_{ij}(\underline{x}))_{i,j=1}^{m}$ 具有如下形式

$$b_{ij}(\underline{x}) = \begin{cases} h_{c,\rho}, & i = j, \\ 0, & i \neq j \ and \ \langle x_i, x_j \rangle \notin \mathcal{I}(\kappa), \\ h_{c,\rho}, & i \neq j \ and \ \langle x_i, x_j \rangle \in \mathcal{I}(\kappa). \end{cases}$$

与之前一样，根据引理*3.1*，我们区分两种情况。

- 首先，假设 κ 不是对称的，因此 $I(\kappa) = \{1\}$。在这种情况下，$B_{2,\mathfrak{S}}(x)$ 是对角且可逆的，因此

$$I_{2;\mathfrak{S}}(\underline{y}; \underline{x}) = \frac{1}{2h_{c,\rho}} \|\underline{y}\|^2. \tag{11}$$

- 现在假设 κ 是对称的，因此 $I(\kappa) = \{-1,1\}$。只要 x 不包含对跖点对，公式 (11) 仍然成立。若其中包含 $2v$ 对径点，即

$$x_{i_1} = -x_{j_1},\ \ldots,\ x_{i_v} = -x_{j_v},$$

令 $\underline{y}_{i_v}$ 和 $\underline{x}_{i_v}$ 表示移除索引为 i_1、……、i_v 的分量后得到的向量。在这种情况下，该问题归结为一个具有 $m - v$ 个点的问题，并且

$$I_{2;\mathfrak{S}}(\underline{y};\underline{x}) = \begin{cases} \dfrac{1}{2h_{c,\rho}} \left\| \underline{y}_{i_v} \right\|^2, & y_{i_1} = y_{j_1},\ \ldots,\ y_{i_v} = y_{j_v}, \\ \infty, & otherwise. \end{cases}$$

3.1 引理 2.1 中数列之间的一种可能联系

尽管原则上可以通过收缩原理由定理 3.2 推导出定理 3.1 中的大偏差原理，但相应的论证将比通过直接且独立的方法建立大偏差原理复杂得多。

我们首先假设 x_1、……、x_m 均不同于北极 N。在这种情况下，数列

$$\{\, T_L(\underline{x}) - T_L(N)\,\mathbf{1}\,\}_{L\geq 1}$$

可以看作数列的连续像

$$\left\{ \left(T_L(\underline{x}) - \frac{a_{00,L}}{\sqrt{\omega_d}}\,\mathbf{1},\, T_L(N) - \frac{a_{00,L}}{\sqrt{\omega_d}} \right) \right\}_{L\geq 1},$$

该数列是定理 3.2 中所考虑数列的略微修改版本。定义连续映射

$$\Psi(z_1, \ldots, z_m, z_{m+1}) := (z_1 - z_{m+1},\ \ldots,\ z_m - z_{m+1}).$$

于是，由收缩原理可得

$$\begin{aligned} &I_{1;\mathfrak{g}}(y_1, \ldots, y_m; \underline{x}) \\ &\qquad = \inf\{ I_{2;\mathfrak{g}}(z_1, \ldots, z_m, z_{m+1};\, \underline{x}, N) : \Psi(z_1, \ldots, z_m, z_{m+1}) = (y_1, \ldots, y_m) \} \\ &\qquad\qquad = \inf_{z_{m+1}\in\mathbb{R}} I_{2;\mathfrak{g}}(y_1 + z_{m+1},\ \ldots,\ y_m + z_{m+1},\ z_{m+1};\, \underline{x}, N), \end{aligned}$$

其中，$\mathfrak{g} = \mathfrak{L}$ 适用于低无序情形，$\mathfrak{g} = \mathfrak{S}$ 适用于稀疏情形。

最后，我们简要讨论这样一种情形：对于某个 $i \in$，$x_i = N\{1, \ldots, m\}$。为避免平凡情形，我们假设 $m \geq 2$。在这种情况下，去掉第 i 个坐标后，可以重复上述论证，从而将问题简化为 $m - 1$ 个点。事实上，只有当 $y_i = 0$ 时，$I_{1;\mathfrak{L}}(y_1, \ldots, y_m;\ x)$ 和 $I_{1;\mathfrak{S}}(y_1, \ldots, y_m;\ x)$ 才是有限的；若 $y_i \neq 0$，则速率函数等于 ∞。

3.2 一些例子

Exa例 3.1（指数激活函数）。在 [16] 作者证明了，对于 r

$$\sigma(x) = e^{-a^2x^2/2},$$

相应的核为

$$\kappa(t) = \sqrt{\frac{2a^2+1}{(a^2+1)^2 - (a^2t)^2}}, \qquad t \in [-1, 1].$$

令 $\gamma := \frac{2a^2+1}{a^4} > 0$，可将其改写为

$$\kappa(t) = \sqrt{\frac{\gamma}{\gamma + 1 - t^2}}, \qquad t \in [-1, 1].$$

直接计算可得 $\kappa'(1) = 1/\gamma$。因此，当 $\gamma > 1$ 时，我们处于低无序区间；当 $\gamma = 1$ 时，处于稀疏区间；当 $\gamma < 1$ 时，处于高无序区间（见例*5.1*）。还需注意，在所有情况下，$I(\kappa) = \{-1,1\}$，因为 κ 是对称的（见引理*3.1*）。

- $\gamma > 1$ **(**低无序区间**)**。 通过对 L 进行归纳可以验证，

$$\kappa_L(t) = \sqrt{\frac{\gamma^{L+1} - \gamma - (\gamma^L - \gamma)t^2}{\gamma^{L+1} - 1 - (\gamma^L - 1)t^2}}$$

  于是，引理 *3.2* 中定义的函数可化简为，

$$\mathfrak{L}(t) := \lim_{L\to\infty} \gamma^L(1 - \kappa_L(t)) = \frac{(\gamma - 1)(1 - t^2)}{2(\gamma - t^2)}, \qquad t \in [-1, 1].$$

- $\gamma = 1$ **(**稀疏区间**)**。 同样通过归纳可得，

$$\kappa_L(t) = \sqrt{\frac{L - (L - 1)t^2}{L + 1 - Lt^2}}.$$

  可以证明， κ_L 在 $t = 1$ 处有两个导数，因此假设 *A* 在 $\rho = 2$ 和 $c = 2$ 下成立。此外，由引理 *3.3* 可得

$$\mathfrak{S}(t) := \lim_{L\to\infty} L\,(1 - \kappa_L(t)) = \begin{cases} 0, & t^2 = 1 \ \textit{(i.e., } t \in \mathcal{I}(\kappa)\textit{)}, \\ \frac{1}{2}, & t^2 < 1 \ \textit{(i.e., } t \notin \mathcal{I}(\kappa)\textit{)}. \end{cases}$$

例3.2 (ReLU激活函数**)**。 在 [10] 中，作者证明了对于

$$\sigma(x) = x \vee 0,$$

相应的核为

$$\kappa(t) = \frac{1}{\pi}\Big(t(\pi - \arccos t) + \sqrt{1 - t^2}\Big), \qquad t \in [-1, 1].$$

特别地， $\kappa'(1) = 1$。该核并非对称，因此 $I(\kappa) = \{1\}$（见引理*3.1*）。与指数情形不同，很难得到 $\kappa_L(t)$的闭式表达式。然而，可以在 $t = 1$ 附近展开 κ：

$$\kappa(t) = t + \frac{1}{\pi}\Big(\sqrt{1 - t^2} - t\arccos t\Big) = t + \frac{2\sqrt{2}}{3\pi}(1 - t)^{3/2} + o\Big((1 - t)^{3/2}\Big) \qquad (t \to 1^-).$$

事实上，

$$\lim_{t\to 1^-} \frac{\sqrt{1 - t^2} - t\arccos t}{(1 - t)^{3/2}} = \frac{2\sqrt{2}}{3},$$

由洛必达法则和经典极限 $\lim_{t\to 1^-} \frac{t}{\sqrt{1-t}\arccos t} = \sqrt{2}$，可知假设 *A* 对$c = \frac{2\sqrt{2}}{3\pi}$ 和 $\rho = \frac{3}{2}$ 成立。

4 函数型结果

本节探讨能否对上述引入的一个（或两个）数列建立函数型大偏差原理；参见定理 3.1 和 3.2。我们的目标是理解上一节得到的有限维结果是否存在有意义的泛函对应结论。我们从低无序区间开始。

定理 4.1。假设 $\kappa'(1) < 1$。那么，数列 $\{(T_L(x) - T_L(N))_{x\in\mathbb{S}^d}\}_{L\geq 1}$ 和 $\left\{\left(T_L(x) - \frac{a_{00}}{\sqrt{\omega_d}}\right)_{x\in\mathbb{S}^d}\right\}_{L\geq 1}$ 满足大偏差原理，具有相同的速度 $(\kappa'(1))^{-L}$，其良好速率函数分别为 I_1 和 I_2，定义如下

$$I_\ell(f) = \sup\Big\{I_{\ell;\mathfrak{L}}\big(f(x_1),\dots,f(x_m);\, x_1,\dots,x_m\big):\ m\geq 1,\ x_1,\dots,x_m\in\mathbb{S}^d,\ x_i\neq x_j \text{ for } i\neq j\Big\}, \tag{12}$$

其中 $\ell = 1,2$。这里，$I_{1;\mathfrak{L}}$ 和 $I_{2;\mathfrak{L}}$ 是定理 *3.1* 和 *3.2* 中定义的速率函数，而 $\mathfrak{L}$ 是引理 *3.2* 中引入的极限函数。

众所周知，I_1 和 I_2 可以用适当的再生核希尔伯特空间中的范数来表示（例如，参见 [6]）。更确切地说，对于 $\ell\in\{1,2\}$，我们有

$$I_\ell(f) = \begin{cases} \dfrac{\|f\|_{b_\ell}^2}{2}, & \text{if } f\in\mathcal{H}(b_\ell),\\ \infty, & \text{otherwise,}\end{cases}$$

其中，$\mathcal{H}(b_\ell)$ 是由协方差核 b_ℓ 生成的再生核希尔伯特空间，其对应的范数为 $\|\cdot\|_{b_\ell}$。这些核定义如下：

$$b_1(z,w) = \mathfrak{L}(\langle z,N\rangle) + \mathfrak{L}(\langle w,N\rangle) - \mathfrak{L}(\langle z,w\rangle),$$

and

$$b_2(z,w) = \frac{\displaystyle\int_{-1}^1 \mathfrak{L}(t)(1-t^2)^{\frac{d}{2}-1}\,dt}{\displaystyle\int_{-1}^1 (1-t^2)^{\frac{d}{2}-1}\,dt} - \mathfrak{L}(\langle z,w\rangle).$$

我们以一条关于重标度场的另一项（函数型）弱收敛性质的注记结束本节。令 $\{(U_L(x))_{x\in\mathbb{S}^d}\}_{L\geq 1}$ 表示定理 4.1 中出现的两个数列之一。

注 4.1. 定理 *4.1* 的证明（见附录 *A.6*）实际上证明了以下过程数列在分布意义下的函数型收敛：

$$\{(\sqrt{v_L}U_L(x))_{x\in\mathbb{S}^d}\}_{L\geq 1}$$

在空间 $C(\mathbb{S}^d)$ 中。特别地，该数列弱收敛到一个中心连续高斯随机场。作为函数型收敛的结果，极限协方差函数必为连续函数。由于极限的有限维分布由协方差矩阵 $B_{1;\mathfrak{L}}(x)$ 和 $B_{2;\mathfrak{L}}(x)$（参见 (7)、(9)）描述，因此可知，上文定义的相应协方差核 b_1 和 b_2 是 $\mathbb{S}^d\times\mathbb{S}^d$ 上的连续函数。特别地，这意味着引理 *3.2* 中所证明存在的函数 $\mathfrak{L}$ 是连续的。

这里给出定理 4.2，它表明，在稀疏情形下，低无序区间内得到的函数型结果（参见定理 4.1）不可能有任何类似的结论。从某种意义上说，这并不出乎意料。事实上，如果族 $\{(\sqrt{v_L}U_L(x))_{x\in\mathbb{S}^d}\}_{L\geq 1}$ 弱收敛到一个中心化的

连续高斯随机场，且其协方差函数连续，那么引理 3.3 中出现的极限协方差结构必然关于空间变量连续。这将迫使函数 $\mathfrak{S}$连续。

然而， $\mathfrak{S}$ 并不连续；因此，在稀疏区间内，既不能成立定理 4.1 中建立的那类泛函大偏差原理，也不能成立注 4.1 中讨论的函数型弱收敛结果。

定理 4.2。 假设 $\kappa'(1) = 1$。则数列 $\{(T_L(x) - T_L(N))_{x\in\mathbb{S}^d}\}_{L\geq 1}$ 和 $\left\{\left(T_L(x) - \frac{a_{00,L}}{\sqrt{\omega_d}}\right)_{x\in\mathbb{S}^d}\right\}_{L\geq 1}$ 不满足速度为 $\overline{\overline{v}}_L = L^{1/(\rho-1)}$ 的大偏差原理。

注 4.2。 定理 *4.2* 表明，这两个数列无法以自然速度 $v_L = L^{1/(\rho-1)}$ 满足大偏差原理。于是可以进一步问，在不同的速度下是否能够成立大偏差原理。在任何更快的速度下，很容易看出这两个数列无法满足大偏差原理。在任何更慢的速度下，有限维分布确实满足大偏差原理，但其对应的是一个退化速率函数。因此，如果这两个数列能够以较慢的速度满足大偏差原理，那么该原理必然由一个退化速率函数支配；在我们看来，这种情形的意义相当有限。关于这一现象的详细讨论，见 [7] 中的命题 *5.8*。

5 关于高无序区间 $\kappa'(1) > 1$的一些评述

本节中我们取 $\kappa'(1) > 1$，并研究数列$\{\kappa_L(t)\}_{L\geq 1}$ 在 $L \to \infty$ 时对每个 $t \in [-1, 1]$ 的渐近行为。实际上，众所周知，情形 $t = 1$可以忽略，因为对于每个 $L \geq 1$，我们都有 $\kappa_L(1) = 1$ 。

值得指出的是，当 $\kappa'(1) > 1$ 时，对于 $t \in [0, 1]$ ，我们有 $\kappa(t) \geq 0$， κ 是递增的（且$\kappa(1) = 1$）；因此存在 $t^* \in [0, 1)$，使得 $\kappa(t^*) = t^*$。于是我们得到如下情形。

- 若 $t \in [0, 1)$，则
$$\lim_{L\to\infty} \kappa_L(t) = t^* \tag{13}$$
因为数列 $\{\kappa_L(t)\}_{L\geq 1}$ 在 $t = t^*$ 时为常数，在 $t > t^*$ 时递减，在 $t < t^*$ 时递增（实际上，当 $t^* = 0$ 时，情形 $t < t^*$ 为空）。
- 若 $t \in (-1, 0)$，则 $\kappa(t) > t$；事实上，由 (3) 可得
$$\kappa(t) - t = \sum_{q\geq 0} \frac{(J_q(\sigma))^2}{q!} t^q - \underbrace{\left(\sum_{q\geq 0} \frac{(J_q(\sigma))^2}{q!}\right)}_{=\kappa(1)=1} t = \sum_{q\geq 0} \frac{(J_q(\sigma))^2}{q!}(t^q - t) > 0,$$
最后一个不等式成立，是因为对于每个 $q \geq 0$，都有 $t^q \geq t$ ；若 $q \neq 1$，则不等式严格成立 $t^q > t$ 。因此，数列 $\{\kappa_L(t)\}_{L\geq 1}$ 递增；又由于在 $(-1, 0)$ 中找不到 κ 的不动点，我们仍可得到 (13)。
- 若 $t = -1$，则当 $\kappa(-1) = -1$ 时，我们有 $\lim_{L\to\infty} \kappa_L(-1) = -1$ ；否则，
$$\lim_{L\to\infty} \kappa_L(-1) = \begin{cases} t^* & \text{if } \kappa(-1) \in (-1, 1) \\ 1 & \text{if } \kappa(-1) = 1. \end{cases}$$

^为完整起见，我们回顾一下 $D_{0,L} \to d_0 < 1$ （因为 $L \to \infty$）；参见 [16] 中的定理 3.3。因此，上文对 $\{\kappa_L(t)\}_{L\geq 1}$ 的渐近分析表明，引理 2.1 中关于零处的两个极限不成立。实际上， [16] 中的定理 3.3 已经针对 $\{T_L(x) - T_L(N)\}_{L\geq 1}$ 陈述了这一点。

可以尝试研究引理 2.1 中数列的函数收敛性。特别地，可以从一大类有限维分布入手，例如对于$x_1, \ldots, x_m \in \mathbb{S}^d$ 满足 $|\langle x_i, x_j\rangle| < 1$ 的 $i \neq j$。

1. 为简单起见，这里我们还假设对于每个 $i \in \{1, \ldots, m\}$都有 $|\langle x_i, N\rangle| < 1$。于是我们有

$$\lim_{L\to\infty} \mathrm{Cov}(T_L(x_i) - T_L(N), T_L(x_j) - T_L(N)) = \begin{cases} 2(1-t^*) & \text{if } i = j \\ 1 - t^* & \text{if } i \neq j. \end{cases}$$

2. 我们有

$$\lim_{L\to\infty} \mathrm{Cov}\left(T_L(x_i) - \frac{a_{00,L}}{\sqrt{\omega_d}}, T_L(x_j) - \frac{a_{00,L}}{\sqrt{\omega_d}}\right) = \begin{cases} 1 - \widehat{d_0} & \text{if } i = j \\ t^* - \widehat{d_0} & \text{if } i \neq j. \end{cases}$$

那么，在这两种情况下，我们都不可能得到函数型弱收敛。事实上，如果这两个数列中的一个弱收敛到某个中心连续高斯随机场，且具有适当的（连续）协方差函数，那么这就与上述关于有限维分布的协方差函数的计算相矛盾。

我们以一个例子作结。

Ex例5.1 (例3.1，其中 $\gamma < 1$)。 可以通过对 L 归纳来验证 t

$$\kappa_L(t) = \sqrt{\frac{\gamma^{L+1} - \gamma - (\gamma^L - \gamma)t^2}{\gamma^{L+1} - 1 - (\gamma^L - 1)t^2}}.$$

那么，容易验证，上面的 t^* 与 $\sqrt{\gamma}$一致，因此 $-$

$$\lim_{L\to\infty} \kappa_L(t) = \begin{cases} \sqrt{\gamma} & \textit{if } t \in (-1, 1) \\ 1 & \textit{if } |t| = 1. \end{cases}$$

A 证明

本节将证明上述所有陈述。

A.1 引理3.1的证明

利用式(3)中的三角不等式，对于每个 $t \in (-1, 1)$，我们得到

$$\Big|\sum_{q\geq 0} \frac{(J_q(\sigma))^2}{q!} t^q\Big| \leq \sum_{q\geq 0} \frac{(J_q(\sigma))^2}{q!} |t|^q < \sum_{q\geq 0} \frac{(J_q(\sigma))^2}{q!} = \kappa(1) = 1, \tag{14}$$

其中严格不等式成立，这是因为 $|t| < 1$，并且至少存在一个指标 $q \geq 1$ 使得$J_q(\sigma) \neq 0$。

由(14)可知，$\mathcal{I}(\kappa) \subseteq \{-1, 1\}$，显然 $1 \in \mathcal{I}(\kappa)$。我们注意到， $\kappa(-1) = 1$ 当且仅当 κ 是偶函数。事实上，由

$$1 = \kappa(1) = \sum_{q\geq 0} \frac{(J_{2q}(\sigma))^2}{(2q)!} + \sum_{q\geq 0} \frac{(J_{2q+1}(\sigma))^2}{(2q+1)!}$$
$$1 = \kappa(-1) = \sum_{q\geq 0} \frac{(J_{2q}(\sigma))^2}{(2q)!} - \sum_{q\geq 0} \frac{(J_{2q+1}(\sigma))^2}{(2q+1)!}$$

对于每个 $q \geq 0$，都有 $J_{2q+1}(\sigma) = 0$。

A.2 引理 3.2 的证明

在证明引理之前，我们先证明，如果 $\kappa'(1) < 1$，那么

$$|\kappa'(t)| < \kappa'(1) \qquad \text{for every } t \in (-1,1). \tag{15}$$

由（3）可知，对于 $t \in (-1,1)$，$\kappa'(t)$ 的导数可以逐项由幂级数求得。此外，根据单调收敛定理，有

$$\kappa'(1) = \lim_{t \to 1^-} \kappa'(t). \tag{16}$$

对 $|\kappa'(t)|$ 的界，完全可以按照（14）中的相同方法得到。

我们来证明该引理的第一部分，即数列

$$\beta_L(t) := \frac{1-\kappa_L(t)}{(\kappa'(1)^L)}$$

在 $[0,\infty)$ 中存在极限。由于对于任意 $L \geq 1$，$\kappa_L(1) = 1$，我们显然得到 $\mathfrak{L}(1) = 0$。令 $t \neq 1$。由于 $\kappa \in C^1([-1,1])$，存在 $\xi_{L,t} \in (t,1) \subset (-1,1)$，使得

$$0 \leq 1-\kappa_L(t) = 1-\kappa(\kappa_{L-1}(t)) = \kappa'(\xi_{L,t})(1-\kappa_{L-1}(t)) \leq \kappa'(1)(1-\kappa_{L-1}(t)),$$

其中不等式由（15）得出。因此，$0 \leq \beta_L(t) \leq \beta_{L-1}(t)$。

因此，我们得到一个单调有界数列，这保证了极限 $\mathfrak{L}(t)$ 的存在（且有限）。

引理第二部分的证明，即 $\mathfrak{L}(t) = 0$ 当且仅当 $t \in \mathcal{I}(\kappa)$，受到 [23] 中定理 1.3.1 的启发。显然，对于 $t \in \mathcal{I}(\kappa)$，有 $\mathfrak{L}(t) = 0$。下文中，我们假设$t \notin \mathcal{I}(\kappa)$。注意，对于 $L \geq 1$

$$\beta_L(t) = (1-t)\prod_{\ell=0}^{L-1}\left(1+\frac{p_\ell(t)}{\kappa'(1)}\right) = (1-t)\exp\left(\sum_{\ell=0}^{L-1}\log\left(1+\frac{p_\ell(t)}{\kappa'(1)}\right)\right) \tag{17}$$

其中

$$p_\ell(t) := \frac{1-\kappa_{\ell+1}(t)}{1-\kappa_\ell(t)} - \kappa'(1) \quad \text{for } \ell \geq 0,$$

其中 $\kappa_0(t) = t$，并且注意到，由于 $p_\ell(t) \geq -\kappa'(1)$，等式 (17) 中的第二个等号有定义。若能证明级数 $\sum_{\ell=0}^{\infty}\log\left(1+\frac{p_\ell}{\kappa'(t)\,(1)}\right)$ 收敛，则证明完成。 ——

根据拉格朗日定理，存在 $\xi_{\ell,t} \in (\kappa_\ell(t),1)$，使得

$$p_\ell(t) = \frac{1-\kappa_{\ell+1}(t)}{1-\kappa_\ell(t)} - \kappa'(1) = \frac{1-\kappa(\kappa_\ell(t))}{1-\kappa_\ell(t)} - \kappa'(1) = \kappa'(\xi_{\ell,t}) - \kappa'(1).$$

由于，当 $\ell \to +\infty$, $\kappa_\ell(t) \to 1$ 时，我们有 $\xi_{\ell,\ t} \to 1$，因此 $\kappa'(\xi_{\ell,\ t}) \to \kappa'(1)$。于是，对每个 $\eta > 0$, – –，都存在 ℓ，使得对每个 $\ell > \ell$,

$$\left|\log\left(1+\frac{p_\ell(t)}{\kappa'(1)}\right)\right| \leq (1+\eta)\frac{|p_\ell(t)|}{\kappa'(1)},$$

并且只需证 明 $\sum_{L=1}^{\infty}|p_L(t)| < \infty$。根据假设 A，当 ℓ 足够大时，我们有

$$p_\ell(t) = \frac{1-\kappa(\kappa_\ell(t))}{1-\kappa_\ell(t)} - \kappa'(1) = -c(1-\kappa_\ell(t))^{\rho-1} + o((1-\kappa_\ell(t))^{\rho-1});$$

then

$$|p_\ell(t)| \leq 2|c|(1-\kappa_\ell(t))^{\rho-1}.$$

最后，我们有

$$(1-\kappa_\ell(t))^{\rho-1} \leq \{\kappa'(1)(1-\kappa_{\ell-1}(t))\}^{\rho-1} \leq \cdots \leq (\kappa'(1)^{\rho-1})^{\ell-\bar{\ell}}(1-\kappa_{\bar{\ell}}(t))^{\rho-1};$$

总之，至少当 ℓ 足够大时，$|p_\ell(t)|$ 可由收敛几何级数中的一项（乘以一个常数）控制，至此证明完成。

A.3 引理 3.3 的证明

首先注意到，根据 $\kappa'(1)=1$，(5) 得出

$$\kappa(t)=t+c(1-t)^{\rho}+o((1-t)^{\rho}) \quad \text{as } t\to 1^{-} \tag{18}$$

对于 $c\neq 0$ 和 $\rho>1$。我们来证明 $c>0$。为此，我们证明对于 $|t|<1$，$\kappa(t)>t$。考虑函数

$$g(t):=\kappa(t)-t;$$

于是我们有 $g'(t)=\kappa'(t)-1$。注意，若 $\kappa'(1)=1$，则 (16) 仍成立，因此 $g'(t)\leq \kappa'(1)-1\leq 0$，并且当 $|t|<1$ 时该不等式为严格不等式；所以 g 是严格递减函数。因此，对于 $|t|<1$，有 $g(t)>g(1)$，即 $\kappa(t)-t>\kappa(1)-1=0$，或等价地，$\kappa(t)>t$，于是 $c>0$。

现在我们讨论引理陈述中的极限；若 $t\in\mathcal{I}(\kappa)$，该极限显然成立；因此，从现在起，在本证明中，我们取 $t\notin\mathcal{I}(\kappa)$。由 (18)（并且由于当 $L\to\infty$ 时，$\kappa_L(t)\to 1$），我们有

$$\begin{aligned}1-\kappa_L(t)&=1-\kappa(\kappa_{L-1}(t))=1-\{\kappa_{L-1}(t)+c(1-\kappa_{L-1}(t))^{\rho}+o((1-\kappa_{L-1}(t))^{\rho})\}\\&=1-\kappa_{L-1}(t)-c(1-\kappa_{L-1}(t))^{\rho}+o((1-\kappa_{L-1}(t))^{\rho})\\&=(1-\kappa_{L-1}(t))\{1-c(1-\kappa_{L-1}(t))^{\rho-1}+o((1-\kappa_{L-1}(t))^{\rho-1})\}\end{aligned}$$

因此（回顾一下，$(1+z)^{-\rho+1}\sim 1+(-\rho+1)z$ 当 $z\to 0$ 时）

$$(1-\kappa_L(t))^{-\rho+1}=(1-\kappa_{L-1}(t))^{-\rho+1}+c(\rho-1)+o(1).$$

Then

$$\lim_{L\to\infty}(1-\kappa_L(t))^{-\rho+1}-(1-\kappa_{L-1}(t))^{-\rho+1}=c(\rho-1)$$

并且，根据施托尔茨–切萨罗定理,

$$\lim_{L\to\infty}\frac{(1-\kappa_L(t))^{-\rho+1}}{L}=c(\rho-1),$$

或者等价地

$$\lim_{L\to\infty}\frac{1-\kappa_L(t)}{L^{-1/(\rho-1)}}=(c(\rho-1))^{-1/(\rho-1)}.$$

证明至此完成。

A.4 定理 3.1 的证明

在此证明中，我们令

$$U_L(\underline{x})=T_L(\underline{x})-T_L(N)\mathbf{1}.$$

因此，由于 $U_L(\underline{x})$ 是一个均值为零、协方差为

$$\mathrm{Cov}(U_L(x_i),U_L(x_j))=\kappa_L(\langle x_i,x_j\rangle)-\kappa_L(\langle x_i,N\rangle)-\kappa_L(\langle x_j,N\rangle)+1,$$

对于每个 $\underline{\theta}\in\mathbb{R}^m$，我们有

$$\frac{1}{v_L}\log\mathbb{E}\left[e^{v_L\langle\underline{\theta},U_L(\underline{x})\rangle}\right]=\frac{v_L}{2}\sum_{i,j=1}^{m}\{\kappa_L(\langle x_i,x_j\rangle)-\kappa_L(\langle x_i,N\rangle)-\kappa_L(\langle x_j,N\rangle)+1\}\theta_i\theta_j.$$

现在，对低无序情形使用引理 3.2，对稀疏情形使用引理 3.3，我们有

$$\lim_{L\to+\infty}v_L(\kappa_L(\langle x_i,x_j\rangle)-\kappa_L(\langle x_i,N\rangle)-\kappa_L(\langle x_j,N\rangle)+1)=\mathfrak{g}(\langle x_i,N\rangle)+\mathfrak{g}(\langle x_j,N\rangle)-\mathfrak{g}(\langle x_i,x_j\rangle)$$

其中 v_L 和 $\mathfrak{g}$ 如定理陈述所示。因此,

$$\lim_{L\to\infty}\frac{1}{v_L}\log\mathbb{E}\left[e^{v_L\langle\underline{\theta},U_L(\underline{x})\rangle}\right]=\frac{1}{2}\langle\underline{\theta},B_{1;\mathfrak{g}}(\underline{x})\underline{\theta}\rangle,$$

其中 $B_{1;\mathfrak{g}}(\underline{x})$ 是式（7）中的方阵。应用 G¨artner Ellis 定理即可得到所需的大偏差原理。

A.5 定理 3.2 的证明

在定理的证明中，我们需要交换极限与积分的顺序。以下引理为这一步提供了依据，其证明推迟到本节末尾。

引理 A.1. 设 $\kappa'(1) \leq 1$。那么

$$\lim_{L\to+\infty} \frac{\int_{-1}^{1} v_L(1-\kappa_L(t))(1-t^2)^{\frac{d}{2}-1}dt}{\int_{-1}^{1}(1-t^2)^{\frac{d}{2}-1}dt} = \frac{\int_{-1}^{1} \mathfrak{g}(t)(1-t^2)^{\frac{d}{2}-1}dt}{\int_{-1}^{1}(1-t^2)^{\frac{d}{2}-1}dt}$$

其中 v_L 和 $\mathfrak{g}$ 与定理 *3.1* 中的相同。

在此证明中，我们令

$$U_L(\underline{x}) = T_L(\underline{x}) - \frac{a_{00,L}}{\sqrt{\omega_d}}\mathbf{1}.$$

由于数列 $\{a_{\ell,m}\}^{(\ell,m)\in\mathcal{A}}$ 是相互独立、均值为零的高斯随机变量，方差为 $\mathbb{E}[a_{\ell m,L}^2] = C_{\ell,L}$，因此我们有

$$\mathrm{Cov}\left(T_L(x_i), \frac{a_{00,L}}{\sqrt{\omega_d}}\right) = D_{0,L}$$

因此

$$\mathrm{Cov}(U_L(x_i), U_L(x_j)) = \kappa_L(\langle x_i, x_j\rangle) - D_{0,L}.$$

因此，对于每个 $\underline{\theta} = (\theta_1、\ldots\ldots、\theta_m) \in \mathbb{R}^m$，都有

$$\frac{1}{v_L}\log\mathbb{E}\left[e^{v_L\langle\underline{\theta},U_L(\underline{x})\rangle}\right] = \frac{v_L}{2}\sum_{i,j=1}^{m}\{\kappa_L(\langle x_i,x_j\rangle) - D_{0,L}\}\theta_i\theta_j\,.$$

Now,

$$v_L\big(\kappa_L(\langle x_i,x_j\rangle) - D_{0,L}\big) = v_L(1-D_{0,L}) - v_L(1-\kappa_L(\langle x_i,x_j\rangle))\,.$$

利用 Gegenbauer 多项式的正交性，我们有

$$D_{0,L} = \frac{\int_{-1}^{1}\kappa_L(t)(1-t^2)^{d/2-1}\mathrm{d}t}{\int_{-1}^{1}(1-t^2)^{d/2-1}\mathrm{d}t}$$

因此

$$v_L(1-D_{0,L}) = \frac{\int_{-1}^{1}(1-\kappa_L(t))(1-t^2)^{d/2-1}\mathrm{d}t}{\int_{-1}^{1}(1-t^2)^{d/2-1}\mathrm{d}t}\,.$$

因此，将引理 A.1 与低无序情形下的引理 3.2 以及稀疏情形下的引理 3.3 结合起来，我们得到

$$\lim_{L\to+\infty} v_L\big(\kappa_L(\langle x_i,x_j\rangle - D_{0,L}\big) = \frac{\int_{-1}^{1}\mathfrak{g}(t)(1-t^2)^{\frac{d}{2}-1}dt}{\int_{-1}^{1}(1-t^2)^{\frac{d}{2}-1}dt} - \mathfrak{g}(\langle x_i,x_j\rangle).$$

然后，应用 G¨artner–Ellis 定理即可得到所需的大偏差原理。

我们将引理 A.1 的证明分为两种情形。

低无序情形下引理 *A.1* 的证明。利用引理 3.2，我们有

$$\beta_L(t) = \frac{1-\kappa_L(t)}{(\kappa'(1))^L} \to \mathfrak{L}(t)\ .$$

由于对每个 $t \in (-1,\ 1)$，都有 $|\kappa'(t)| < 1$，利用 L 次中值定理，我们有

$$0 \leq \beta_L(t) \leq 1-t \tag{19}$$

因此 $\beta_L \leq 2$，结论由控制收敛定理得出。 □

稀疏情形下引理 *A.1* 的证明。与前一种情形一样，使用引理 3.3

$$\eta_L(t) := L^{1/(\rho-1)}(1-\kappa_L(t)) \to \mathfrak{S}(t).$$

为了应用控制收敛定理，我们证明存在一个常数 $M>0$，使得 $0 \le \eta_L(t) \le M$。对每个 $L \ge 1$，令

$$R_L := \{t \in [-1,1] : \kappa_L(t) > 0\}.$$

根据构造，$\{R_L\}_{L\ge 1}$ 是 $[-1,1]$ 的一个开覆盖；事实上，对每个 $t \in [-1,1]$，都有 $\kappa_L(t) \to 1$。于是，由于 $[-1,1]$ 是一个紧集，存在 $[-1,1]$ 的有限子覆盖；因此，对某个 L，我们有

$$[-1,1] = \bigcup_{i=1}^{\bar{L}} R_i$$

并且由于当 $t \in (0,1]$ 时，我们有 $\kappa(t) > 0$，因此有 $R_L \subset R_{L+1}$。故 $[-1,1] = R_{\bar{L}}$。现在我们取 $L > L$，并且对每个 $t \in [-1,1]$，得到

$$\kappa_L(t) = \kappa_{L-\bar{L}}(\kappa_{\bar{L}}(t)) \ge \kappa_{L-\bar{L}}(0),$$

由此得到

$$0 \le L^{1/(\rho-1)}(1-\kappa_L(t)) \le \frac{L^{1/(\rho-1)}}{(L-\bar{L})^{1/(\rho-1)}}(L-\bar{L})^{1/(\rho-1)}(1-\kappa_{L-\bar{L}}(0)),$$

并且由引理 3.3，上界收敛到 $\mathfrak{S}(0)$（当 $L \to \infty$ 时）。总之，证明已完成，因为收敛数列有界。□

A.6 定理 4.1 的证明

在整个证明中，我们用 $\{(U_L(x))_{x\in S^d}\}_{L\ge 1}$ 表示定理陈述中的两个数列之一。在这两种情况下，有限维分布都满足大偏差原理（见定理 3.1 和 3.2）。根据 Dawson-Gärtner 定理，数列 $\{(U_L(x))_{x\in S^d}\}_{L\ge 1}$ 在 $C(\mathbb{S}^d)$ 中的任何大偏差原理（如果存在），其速率函数都必须由（12）给出。根据大偏差原理的经典结果，要证明该定理，只需证明数列 $\{(U_L(x))_{x\in\mathbb{S}^d}\}_{L\ge 1}$ 在 $C(\mathbb{S}^d)$ 中指数紧。也就是说，对每个 $M>0$，都存在一个紧集 $K_M \subset C(\mathbb{S}^d)$，使得

$$\limsup_{L\to\infty} \frac{1}{v_L} \log \mathbb{P}((U_L(x))_{x\in\mathbb{S}^d} \notin K_M) \le -M.$$

利用 [6] 中的定理 2.4 证明数列 $\{(U_L(x))_{x\in\mathbb{S}^d}\}_{L\ge 1}$ 以速度 v_L 指数紧，只需证明 $\{(\sqrt{v_L}U_L(x))_{x\in\mathbb{S}^d}\}_{L\ge 1}$ 在 $C(\mathbb{S}^d)$ 中紧；感兴趣的读者可参见 [6]，了解紧性与指数紧性之间关系的更多细节。

由该数列有限维分布的弱收敛可知

$$\{(\sqrt{v_L}\,U_L(x))_{x\in\mathbb{S}^d}\}_{L\ge 1},$$

如果能够证明，存在某些常数 a，$b>0$，且对于每个 x，$y \in \mathbb{S}^d$ 以及每个 $L \ge 1$，

$$\mathbb{E}[|\sqrt{v_L}U_L(x) - \sqrt{v_L}U_L(y)|^a] \le c\,\|x-y\|^{b+d},$$

那么，根据 [22]，定理 23.7，数列 $\{(\sqrt{v_L}\,U_L(x))_{x\in\mathbb{S}^d}\}_{L\ge 1}$ 在 $C(\mathbb{S}^d)$ 中是紧的，且其极限具有连续样本路径。

下面证明，当 $r > 0$ 足够大时，前一个不等式对 $a = 2r$ 和 $b = 2r - d$ 成立。我们注意到，在这两种情况下，

$$U_L(x) - U_L(y) = T_L(x) - T_L(y) \overset{d}{=} \sqrt{2(1 - \kappa_L(\langle x, y\rangle))}\, Z,$$

其中 $\overset{d}{=}$表示依分布相等， Z 是一个标准正态随机变量。因此，

$$\mathbb{E}[|\sqrt{v_L}U_L(x) - \sqrt{v_L}U_L(y)|^{2r}] = (2v_L(1 - \kappa_L(\langle x, y\rangle)))^r\, \mathbb{E}[|Z|^{2r}].$$

现在，由于 $v_L = (\kappa'(1))^{-L}$，回想起 $v_L(1 - \kappa_L(\langle x,\ y\rangle)) = \beta_L(\langle x,\ y\rangle)$，并利用 (19)，我们得到

$$2v_L(1 - \kappa_L(\langle x, y\rangle)) \le 2(1 - \langle x, y\rangle) = \|x - y\|^2,$$

其中我们用到了 $\|x\| = \|y\| = 1$。综上，我们有

$$\mathbb{E}[|\sqrt{v_L}U_L(x) - \sqrt{v_L}U_L(y)|^{2r}] \le \mathbb{E}[|Z|^{2r}]\, \|x - y\|^{2r},$$

由此得到所需的估计。

A.7 定理 4.2 的证明

对于两个数列，我们都用反证法进行论证。若以速度 $v_L = L^{1/(\rho-1)}$成立函数型大偏差原理，类比定理 4.1 中陈述的大偏差原理（显然，此处 $\mathfrak{S}$ 是引理 3.3 中的函数），则应有以下结论。

1. $\{(T_L(x) - T_L(N))_{x\in\mathbb{S}^d}\}_{L\ge1}$ 满足速度为 $v_L = L^{1/(\rho-1)}$、速率函数为 I_1 的大偏差原理，其中 I_1 定义为

$$I_1(f) = \sup\{I_{1;\mathfrak{S}}(f(x_1), \dots, f(x_m); x_1, \dots, x_m) : m \ge 1,\ x_1, \dots, x_m \in \mathbb{S}^d,\ x_i \ne x_j \text{ for } i \ne j\},$$

2. $\left\{\left(T_L(x) - \frac{a_{00}}{\sqrt{\omega_d}}L\right)_{x\in\mathbb{S}^d}\right\}_{L\ge1}$ 满足速度为 $v_L = L^{1/(\rho-1)}$、速率函数为 I_2 的大偏差原理，其中 I_2 定义为

$$I_2(f) = \text{上确界}\{I_{2;\mathfrak{S}}(f(x_1), \dots, f(x_m); x_1, \dots, x_m) : m \ge 1,\ x_1, \dots, x_m \in \mathbb{S}^d,\ x_i \ne x_j \text{ for } i \ne j\}$$

我们将看到，对于一类过于宽泛的函数， I_i 是良好速率函数，而 $I_i(f) = +\infty$ （ $i = 1$ 和 $i = 2$ 两种情形有所区别；见下文）。因此，如果我们尝试通过应用收缩原理来得到有限维速率函数，就会预期

$$I_{i;\mathfrak{S}}(y_1, \dots, y_m; x_1, \dots, x_m) = \inf\{I_i(f) : f(x_1) = y_1, \dots, f(x_m) = y_m\} \tag{20}$$

对于 $x_1, \dots, x_m \in \mathbb{S}^d$ 和 $y_1, \dots, y_m \in \mathbb{R}$ 的每一种取值都是如此。然而，这会导致矛盾，因为上面的下确界通常为 $+\infty$，而实际的有限维速率函数是有限的。

1. 设 $f \in C(\mathbb{S}^d)$ 是一个非常数函数。那么存在一个闭集 $U \subseteq \mathbb{S}^d$，使得 f 在 U 上是非常数函数，且 U 不包含对径点。若设 $m_f := \min\{f(x) : x \in U\}$，且 $M_f := \max\{f(x) : x \in U\}$，则有 $m_f < M_f$。此外，容易证明存在 $\varepsilon > 0$，使得对每个 m，都可以找到不同点 $x_1, \dots, x_m \in \mathbb{S}^d$，满足

$$\frac{1}{m}\sum_{i=1}^{m} f(x_i) = \frac{m_f + M_f}{2}$$

and

$$\left|f(x_1)-\frac{m_f+M_f}{2}\right|,\ldots,\left|f(x_m)-\frac{m_f+M_f}{2}\right|>\varepsilon.$$

（取 m 为偶函数即可轻松做到，但这并非必要；此外，这一假设不会限制证明。）然后，考虑定理 3.2 中速率函数的注3.3所述讨论，可得

$$\begin{aligned}
I_1(f) &\geq \frac{1}{2(m+1)h_{c,\rho}}\left\{m\sum_{i=1}^{m}(f(x_i))^2-2\sum_{1\leq i<j\leq m}f(x_i)f(x_j)\right\}\\
&= \frac{1}{2(m+1)h_{c,\rho}}\left\{(m+1)\sum_{i=1}^{m}(f(x_i))^2-\left(\sum_{i=1}^{m}f(x_i)\right)^2\right\}\\
&= \frac{1}{2h_{c,\rho}}\left\{\sum_{i=1}^{m}(f(x_i))^2-\frac{1}{m+1}\left(\sum_{i=1}^{m}f(x_i)\right)^2\right\}\\
&\geq \frac{1}{2h_{c,\rho}}\left\{\sum_{i=1}^{m}(f(x_i))^2-\frac{1}{m}\left(\sum_{i=1}^{m}f(x_i)\right)^2\right\},
\end{aligned}$$

因此

$$I_1(f)\geq\frac{1}{2h_{c,\rho}}\sum_{i=1}^{m}\left(f(x_i)-\frac{1}{m}\sum_{j=1}^{m}f(x_j)\right)^2\geq\frac{m\varepsilon^2}{2h_{c,\rho}}.$$

鉴于上述构造对每个 $m\in\mathbb{N}$ 都是可能的，可以推出 $I_1(f)=+\infty$。此外，为了应用收缩原理，我们需要验证 I_1 是一个良好速率函数。特别是，我们必须证明每个水平集

$$\{f:I_1(f)\leq\eta\},\quad \eta\geq 0,$$

约化为常值函数的集合是一个紧集。假设存在 $c\in\mathbb{R}$，使得对于所有 $x\in\mathbb{S}^d$，$f(x)=c$。那么，如果我们将注意力限制在满足 $\langle x_i,x_j\rangle\notin\mathcal{I}(\kappa)$（对每个 $i,j\in\{1,\ldots,m\}$）的不同点 $x_1,\ldots,x_m\in\mathbb{S}^d$ 上，我们有

$$I_1(f)\geq\sup_{m\geq 1}\frac{1}{2(m+1)h_{c,\rho}}\left(m^2c^2-2\frac{m(m-1)}{2}c^2\right)=\sup_{m\geq 1}\frac{mc^2}{2(m+1)h_{c,\rho}}=\frac{c^2}{2h_{c,\rho}}.$$

因此，该水平集已知是闭集；由于它是 $\left\{c\in\mathbb{R}\text{ 的子集：}\ \frac{c^2}{2h_{c,\rho}}\leq\eta\right\}$（其有界），所以它也是有界的；于是我们可以说 I_1 是一个良好速率函数。总之，取 $y_1,\ldots,y_m$ 且不全相同，(20) 不成立。

2. 取 $f\in C(\mathbb{S}^d)$，使得对于某个 $x\in\mathbb{S}^d$，有 $f(x)\neq 0$。那么存在 $\varepsilon>0$，使得对于每个 m，我们都可以找到 $x_1,\ldots,x_m\in\mathbb{S}^d$，满足

$$|f(x_1)|,\ldots,|f(x_m)|>\varepsilon.$$

考虑到注3.4中关于定理 3.2 的速率函数的讨论（将范围限制为不同点 $x_1,\ldots,x_m$，使得对每个 $i,\ j$，都有 $\langle x_i,\ x_j\rangle\notin\mathcal{I}(\kappa)$；鉴于我们将要证明的内容，这不会造成任何问题），我们有

$$I_2(f)\geq\frac{1}{2h_{c,\rho}}\sum_{i=1}^{m}(f(x_i))^2\geq\frac{m\varepsilon^2}{2h_{c,\rho}}.$$

由于对每个 $m \in \mathbb{N}$ 都可以进行这种构造，因此可得 $I_2(f) = +\infty$。此外，若对所有 $x \in \mathbb{S}^d$，都有 $f(x) = 0$，则 $I_2(f) = 0$（这一点很容易验证），因此

$$I_2(f) = \begin{cases} 0 & \text{if } f(x) = 0 \text{ for all } x \in \mathbb{S}^d, \\ +\infty & \text{otherwise.} \end{cases}$$

因此，对于每个 $\eta \geq 0$，水平集 $\{f : I_2(f) \leq \eta\}$ 都是紧致的，因为它退化为单点（零函数）。因此， I_2 是一个良好速率函数。总之，当 y_1，……， y_m 不全为零时，(20) 不成立。

参考文献

[1] Robert J. Adler 和 Jonathan E. Taylor，《随机场与几何》，施普林格数学专著，施普林格，纽约，2007。MR2319516[2] Luisa Andreis、Federico Bassetti 和 Christian Hirsch，《全连接神经网络中协方差过程的大偏差原理》，2025。预印本。arXiv:2505.08062。[3] Kendall Atkinson 和 Weimin Han，《单位球面上的球谐函数与逼近：导论》，数学讲义，第 2044 卷，施普林格，海德堡，2012。MR2934227[4] Jean-Marc Aza¨ıs 和 Mario Wschebor，《随机过程与随机场的水平集和极值》，John Wiley & Sons, Inc.，新泽西州霍博肯，2009。MR2478201[5] Krishnakumar Balasubramanian、Larry Goldstein、Nathan Ross 和 Adil Salim，《通过 Stein方法逼近高斯随机场及其在宽型随机神经网络中的应用》，应用与计算调和分析 72 (2024)，论文编号 101668，共 27 页。MR4746898[6] Paolo Baldi，《高斯概率的紧性与指数紧性》，ESAIM Probab. Stat. 24 (2020)，113–126。MR4071318[7] Paolo Baldi 和 Barbara Pacchiarotti，《连续高斯过程的大偏差：从小噪声到短时间》，Stochastics (2025)。即将刊出。[8] Alberto Bietti 和 Francis Bach，《在核区域中，深层ReLU网络等同于浅层网络》，国际学习表征会议，2021。https://openreview.net/forum?id=aDjoksTpXOP。[9] Valentina Cammarota、Domenico Marinucci、Michele Salvi 和 Stefano Vigogna，《浅层神经网络的定量泛函中心极限定理》，Mod. Stoch. Theory Appl. 11 (2024)，第 1 期，85–108。MR4692019[10] Youngmin Cho 和 Lawrence K. Saul，《用于大间隔分类的反余弦核分析与扩展》，2011。预印本。arXiv:1112.3712。[11] George Cybenko，《通过S形函数的叠加进行逼近》，Math. Control Signals Systems 2 (1989)，第 4 期，303–314。MR1015670[12] Amit Daniely、Roy Frostig 和 Yoram Singer，《深入理解神经网络：初始化的力量与表达能力的对偶视角》，神经信息处理系统进展，2016。[13] Amir Dembo 和 Ofer Zeitouni，《大偏差技术及其应用》，第 2 版，数学应用（纽约），第 38 卷，施普林格出版社，纽约，1998。MR1619036[14] Simmaco Di Lillo，《随机神经网络的临界点》，2025。预印本。arXiv:2505.17000。[15] Simmaco Di Lillo、Domenico Marinucci、Michele Salvi 和 Stefano Vigogna，《深度神经网络的分形几何与正则几何》，2025。预印本。arXiv:2504.06250。[16] ,《深度神经网络的谱复杂度》，SIAM J. Math. Data Sci. 7 (2025)，第 3 期，1154–1183。MR4941937[17] Stefano Favaro、Boris Hanin、Domenico Marinucci、Ivan Nourdin 和 Giovanni Peccati，《深度神经网络中的定量中心极限定理》，Probab. Theory Related Fields 191 (2025)，第 3-4 期，933–977。MR4898096[18] Boris Hanin，《无限宽极限下作为高斯过程的随机神经网络》，Ann. Appl. Probab. 33 (2023)，第 6A 期，4798–4819。MR4674064[19] Christian Hirsch 和 Daniel Willhalm，《单隐藏层神经网络的大偏差》，Stoch. Dyn. 24 (2024)，第 8 期，论文编号 2550002，共 53 页。MR4873338[20] Kurt Hornik、Maxwell Stinchcombe 和 Halbert White，《多层前馈网络是通用逼近器》，Neural Networks 2 (1989)，359–366。

[21] Arthur Jacot、Franck Gabriel 和 Cl´ement Hongler，神经切核，《神经信息处理系统进展》，2018。[22] Olav Kallenberg，现代概率论基础，第 3 版，《概率论与随机建模》，第 99 卷，施普林格，Cham，[2021] ©2021。MR4226142[23] Marek Kuczma、Bogdan Choczewski 和 Roman Ger，迭代函数方程，《数学及其应用百科全书》，第 32 卷，剑桥大学出版社，剑桥，1990。MR1067720[24] Jaehoon Lee、Jascha Sohl-Dickstein、Jeffrey Pennington、Roman Novak、Samuel S. Schoenholz 和 Yasaman Bahri，深度神经网络作为高斯过程，国际学习表征会议，2018。https://openreview.net/forum?id=B1EA-M-0Z。[25] Claudio Macci、Barbara Pacchiarotti 和 Giovanni Luca Torrisi，高斯神经网络的大偏差与中等偏差，《应用概率学杂志》(2026)。即将刊出。[26] Domenico Marinucci 和 Giovanni Peccati，球面上的随机场，伦敦数学学会讲义丛书，第 389 卷，剑桥大学出版社，剑桥，2011。表示、极限定理与宇宙学应用。MR2840154[27] Alexander G. de G. Matthews、Jiri Hron、Mark Rowland、Richard E. Turner 和 Zoubin Ghahramani，宽深度神经网络中的高斯过程行为，国际学习表征会议，2018。https://openreview.net/forum?id=H1-nGgWC-。[28] Radford M. Neal，无限网络的先验分布，《神经网络的贝叶斯学习》，1996，第 29–53 页。[29] Carl Edward Rasmussen 和 Christopher K. I. Williams，机器学习中的高斯过程，《自适应计算与机器学习》，麻省理工学院出版社，马萨诸塞州剑桥，2006。MR2514435[30] G´abor Szeg¨o，正交多项式，第 4 版，《美国数学会讨论会出版物》，第 XXIII 卷，美国数学会，罗得岛州普罗维登斯，1975。MR372517[31] Quirin Vogel，具有*ReLU*激活的高斯神经网络的大偏差，《统计与概率快报》**230** (2026)，论文编号 110611。MR4994611[32] Christopher Williams，无限网络计算，《神经信息处理系统进展》，1996。[33] Myhailo Yadrenko，随机场谱理论，《数学与工程翻译系列》，Optimization Software, Inc.，出版部，纽约，1983。俄文译本。MR697386