arXiv:2304.09398v2 [math.ST] 2 Oct 2024

稀疏加性模型中的极小极大信号检测

Subhodh Kotekal 和
ChaoGao 芝加哥大学

摘要

在需要兼顾高维性与建模灵活性的情况下，稀疏加性模型是一种颇具吸引力的选择。我们研究信号检测问题，并建立了检测稀疏加性信号的极小极大分离率。我们的结果是非渐近的，适用于单变量分量函数属于一般再生核希尔伯特空间的情形。与估计理论不同，极小极大分离率揭示了稀疏性与函数空间选择之间非平凡的相互作用。我们还研究了对稀疏性的自适应，并为一般函数空间建立了自适应检验率；在某些空间中可以实现自适应，而另一些空间则带来不可避免的代价。最后，我们在索об列夫空间的设定下研究了对稀疏性和光滑性的同时自适应，并纠正了文献中的一些既有论断。

1 引言

为提高可解释性、计算效率，并规避困扰高维回归的维数灾难，人们常对真实回归函数作出结构假设。事实上，稀疏线性回归或许可以说是现代统计学的标志性成果。尽管稀疏线性回归广受欢迎，但在某些情况下，为了捕捉协变量的非线性效应，可能需要更大的灵活性。Hastie 和 Tibshirani 力求在灵活性与结构之间取得平衡，[20] 提出了广义加性模型（GAM），作为备受推崇的线性模型的一种自然扩展。在 GAM 中，回归函数可表示为若干单变量（非线性）分量函数的加性分解。然而，与线性模型一样，要实现相合估计，样本量必须大于维数。遵循现代统计学的直觉，稀疏加性模型颇具吸引力，[29, 33, 35, 38, 39, 50]。其回归函数可表示为若干单变量函数的加性分解，其中只有一个较小的子集非零；它结合了 GAM 与稀疏性。

为明确记号，考虑 p维高斯白噪声模型

$$dY_x = f(x)\,dx + \frac{1}{\sqrt{n}} dB_x \tag{1}$$

对于 $x \in [0,1]^p$。尽管考虑非参数回归模型 $Y_i = f(X_i) + \epsilon_i$ （例如 [38] 中的模型）可能更贴近实际数据分析，但高斯白噪声模型很方便，因为它避免了令人分心的技术细节，同时使我们能够专注于统计学本质。事实上，非参数统计学文献长期以来一直通过研究白噪声模型来理解理论极限，并依赖众所周知的渐近等价 [5, 40] ；这些等价关系表明，在各种条件下，一个模型中获得的数学结果可以迁移到另一个模型中。由于我们关注的是理论而非实践，因此沿袭这一传统。广义加性模型认为，回归函数具有如下形式： $f(x) = \sum_{j=1}^{p} f_j(x_j)$，其中 f_1、...、 f_p 是属于某个函数空间 $\mathcal{H}$ 的单变量函数。同样，稀疏加性模型认为

$$f(x) = \sum_{j \in S} f_j(x_j)$$

存在某个未知的支撑集 S，其大小为 s，表示活跃协变量。

现有文献大多研究稀疏加性模型的估计，主要聚焦于非参数回归情形，并令 $\mathcal{H}$ 为再生核希尔伯特空间。经过一系列研究，[29, 33, 35, 39], Raskutti 等人 [38]（另见 [29]）证明，在对再生核希尔伯特空间 $\mathcal{H}$ 采用不同选择时，惩罚型 M 估计量可达到极小极大估计率。Yuan 和 Zhou [50] 在近似稀疏性的概念下建立了极小极大估计率。如今，经验过程这一强大框架已成为估计理论中的典型工具，并被用于他们的证明中。还有一些文章研究了稀疏加性模型的推广。例如，[49] 的作者考察了包括加性信号在内的若干结构 $f(x) = \sum_{j=1}^{k} f_j(x)$，其中每个分量函数 f_j 实际上是一个至多依赖 s_j 个坐标的多元函数，并且是 α_j-H¨older 函数。作者进一步推导出极小极大速率，以处理坐标函数光滑性指标和稀疏性的异质性；作为其结果的一个例子，他们表明，在某一特定情形及一些条件下，对于 $s_j = s$ 且所有 j 均满足 $\alpha_j = \alpha$ 的特殊齐次情形，速率为 $kn^{-\frac{2\alpha}{2\alpha+s}} + ks\log\left(\frac{p}{s}\right)/n$。近期，[3] 的研究结果表明，某些深度神经网络能够在稀疏 k 阶交互模型中达到极小极大估计率。k 阶交互模型也称为非参数方差分析。举例来说，稀疏二阶交互模型假设 $f(x) = \sum_{j\in S_1} f_j(x_j) + \sum_{(k,l)\in S_2} f_{kl}(x_k, x_l)$，其中活跃变量集合 S_1 和交互项集合 S_2 的基数都很小。当 f_j 为 β_1-H¨older 函数且 f_{kl} 为 β_2-H¨older 函数时，[3] 在一些条件下，并忽略关于 n 的对数因子后，得到速率 $s_1(n^{-\frac{2\beta_1}{2\beta_1+1}} + (\log p)/n) + s_2(n^{-\frac{2\beta_2}{2\beta_2+2}} + (\log\ p)/n)$。

关于信号检测问题，文献中的讨论要少得多

$$H_0 : f \equiv 0, \tag{2}$$

$$H_1 : ||f||_2 \geq \varepsilon \text{ and } f \in \mathcal{F}_s \tag{3}$$

其中 $\mathcal{F}_s$ 是由（4）给出的稀疏加性信号类。采用极小极大视角 [6,21–24], 目标是确定最小速率 ε，使其作为稀疏度 s、维数 p、样本量 n和函数空间 $\mathcal{H}$ 的函数时，可以对备择假设相对于零假设进行一致检测。

尽管相关估计理论对此讨论得多得多，但在稀疏加性模型之外的各种高维情形中，已经建立了最优检验速率。最经典的设定——高斯序列模型——或许是研究最多的 [2, 7, 8, 10, 12, 14, 18, 24, 30, 34, 41]。在线性回归 [1, 26, 37] 及其他模型 [13, 36]中，也已建立最优速率。一个常见的特点是，最优检验速率呈现出不同于最优估计速率的现象。

回到 (2)-(3)，文献中唯一直接相关的研究是 Ingster 和 Lepski 的文章 [25]，以及 Gayraud 和 Ingster 后来发表的文章 [17]。Ingster 和 Lepski [25] 研究了一种稀疏多通道模型；经过转换到序列空间后，该模型与 (2)-(3) 密切相关。他们采用渐近设定，并且只选择 $\mathcal{H}$ 为 Sobolev空间，由此建立了渐近极小极大分离率。然而，他们的结果仅涉及常数 $\delta \in (0, 1/2)$ 下的情形 $p = O(s^2)$ 和 $s = O(p^{1/2-\delta})$。他们的结果未能精确确定相变 $s \asymp \sqrt{p}$附近的分离率。本质上，他们在稀疏区域的检验程序是对一组 χ^2型检验应用 Bonferroni校正，每个 p 坐标对应一个检验。因此，他们的分离率在 $\sqrt{p}$ 附近存在缺口并不意外。在稠密区域，他们使用单个 χ^2型检验，这与极小极大检验文献中的常见做法一致。Ingster 和 Lepski [25] 还研究了对稀疏性的自适应，以及同时对稀疏性和光滑性的自适应。Ingster 和 Gayraud [17] 研究的是稀疏加性模型，而非稀疏多通道模型，但同样选择了 $\mathcal{H}$，并采用渐近设置。他们通过一种高阶批评型检验统计量，为稀疏情形 $s = p^{1/2-\delta}$ 建立了尖锐常数。本文中，我们主要将结果与 [25] 进行比较，因为它是首篇建立分离率的文章。我们的结果并未涉及尖锐常数的问题。

本文的主要贡献如下。首先，采用 [2], 开创的非渐近极小极大检验框架，我们确立了对于（2）-（3），在稀疏性 s、维数 p、样本量 n以及一般函数空间 $\mathcal{H}$的任意配置下的非渐近极小极大分离率。值得注意的是，我们并未像 [17, 25]那样将范围限定于索об列夫（或 Besov）空间。我们分析的检验程序涉及阈值化的 χ^2 统计量，沿用了其他稀疏信号检测问题中采用的策略 [8–10, 30, 34]。

我们的第二项贡献是为一般函数空间建立自适应检验率。通常，稀疏度是未知的，因此，拥有一种能够适应一般 $\mathcal{H}$ 的方法在实践中具有重要意义。有趣的是，对 $\mathcal{H}$ 的某些选择而言，并不存在任何自适应代价；也就是说，无需知道稀疏度即可达到极小极大速率。事实证明，我们的速率所产生的自适应代价是 $\mathcal{H}$ 的一个精细函数，从而拓展了 Ingster 和 Lepski 对 Sobolev空间 [25] 的研究。在 Sobolev 情形下，我们的结果甚至进一步拓展了他们的文章；在接近情形 $s \asymp \sqrt{p}$ 时，我们的检验通过仅产生涉及 $\log\log\log p$ 的代价，而非其检验所产生的涉及 $\log\log p$ 的代价，提供了更精细的结果。原则上，我们的结果可用于反向寻找一个空间 $\mathcal{H}$，使得该自适应检验率产生给定的自适应代价。

最后，我们在 Sobolev空间的背景下研究了对稀疏性和光滑性的同时自适应。我们发现并纠正了 [25] 所提出的一项论断中的问题。

1.1 符号

本文将始终使用以下符号。对于 $p \in \mathbb{N}$，令 $[p] := \{1, ..., p\}$。对于 $a, b \in \mathbb{R}$，记 $a \vee b := \max\{a, b\}$ and $a \wedge b = \min\{a, b\}$。记 $a \lesssim b$表示存在一个通用常数 $C > 0$，使得 $a \leq Cb$。表达式 $a \gtrsim b$ 表示 $b \lesssim a$。此外， $a \asymp b$ 表示 $a \lesssim b$and $b \lesssim a$。符号 $\langle \cdot, \cdot \rangle$ 表示欧几里得空间中的通常内积，而 $\langle \cdot, \cdot \rangle_F$表示弗罗贝尼乌斯内积。可测空间 $(\mathcal{X}, \mathcal{A})$ 上两个概率测度 P and$Q$ 之间的全变差距离定义为 $d_{TV}(P, Q) := \sup_{A \in \mathcal{A}} |P(A) - Q(A)|$。如果 Q 关于 P 绝对连续，则 χ^2-散度定义为 $\chi^2(Q \,||\, P) := \int_{\mathcal{X}} \left(\frac{dQ}{dP} - 1\right)^2 dP$。对于有限集 S，记号 $|S|$ 表示 S 的基数。全文将使用迭代对数（例如 $\log\log\log p$ 和 $\log\log(np)$ 这样的表达式）。若不特别说明，当此类表达式小于 1 时，我们将其视为某个通用常数。例如，当 $p < e^{e^e}$ 时，$\log\log\log p$ 应理解为等于某个通用常数。

1.2 设置

1.2.1 再生核希尔伯特空间（RKHS）

遵循相关文献（例如 [47]）， $\mathcal{H}$ 将表示一个再生核希尔伯特空间（RKHS）。在讨论主要结果之前，先回顾再生核希尔伯特空间的一些基本性质 [47]。设 $\mathcal{H} \subset L^2([0, 1])$ 是一个再生核希尔伯特空间（RKHS），其对应的内积为 $\langle \cdot, \cdot \rangle_{\mathcal{H}}$。存在一个对称函数 $K\colon [0, 1] \times [0, 1] \to \mathbb{R}_+$，称为 $kernel$，使得对于任意 $x \in [0, 1]$，我们有 (1) 函数 $K(\cdot, x) \in \mathcal{H}$，并且 (2) 对所有 $f \in \mathcal{H}$，有 $f(x) = \langle f, K(\cdot, x) \rangle_{\mathcal{H}}$。Mercer定理（ [47]中的定理 12.20）保证核函数 K 可以按特征函数 $\{\psi_k\}_{k=1}^{\infty}$展开，即 $K(x, x') = \sum_{k=1}^{\infty} \mu_k \psi_k(x) \psi_k(x')$。举例来说，核函数 $K(x, x') = 1 + \min\{x, x'\}$定义了一阶索об列夫空间，其特征值衰减为 $\mu_k \asymp k^{-2}$；核函数 $K(x, x') = \exp\left(-\frac{(x-x')^2}{2}\right)$的特征值衰减则为 $\mu_k \asymp e^{-ck \log^k}$（更详细的综述见 [47]）。

不失一般性，我们将特征值按 $\mu_1 \geq \mu_2 \geq ... \geq 0$排序。特征函数$\{\psi_k\}_{k=1}^{\infty}$ 在 $L^2([0, 1])$中关于通常的 L^2 内积 $\langle \cdot, \cdot \rangle_{L^2}$ 构成正交归一系，且内积$\langle \cdot, \cdot \rangle_{\mathcal{H}}$ 满足 $\langle f, g \rangle_{\mathcal{H}} = \sum_{k=1}^{\infty} \mu_k^{-1} \langle f, \psi_k \rangle_{L^2} \langle g, \psi_k \rangle_{L^2}$，其中 $f, g \in \mathcal{H}$。

1.2.2 参数空间

本节定义了全文中将使用的参数空间。设$\mathcal{H}$ 是一个再生核希尔伯特空间。回顾一下，我们关注的是对于某个稀疏模式 $f(x) = \sum_{j \in S} f_j(x_j)$的稀疏加性信号

。为方便起见，遵循 [17, 38],，我们假设对于所有 j，$\int_0^1 f_j(t)\,dt = 0$。这一假设较为温和，可以放宽；进一步的讨论见第5.1节。令$\mathbf{1} \in L^2([0,1])$表示恒等于一的常函数，考虑 $\mathcal{H}_0 := \mathcal{H} \cap \text{span}\{\mathbf{1}\}^\perp$ 是 $\mathcal{H}$ 的一个闭子空间。因此，$\mathcal{H}_0$ 也是一个再生核希尔伯特空间。我们将搁置 $\mathcal{H}$ （及其特征函数和特征值），仅研究 $\mathcal{H}_0$。令$\{\psi_k\}_{k=1}^\infty$ 和 $\{\mu_k\}_{k=1}^\infty$ 分别表示其对应的特征函数和特征值。为方便起见，遵循 [2],，我们也假设 $\mu_1 = 1$；这一条件很容易放宽。

对于每个子集 $S \subset [p]$，定义

$$\mathcal{H}_S := \left\{ f(x) = \sum_{j \in S} f_j(x_j) : f_j \in \mathcal{H}_0 \text{ and } ||f_j||_{\mathcal{H}_0} \leq 1 \text{ for all } j \in S \right\}.$$

再生核希尔伯特空间范数上的条件强制了正则性。定义参数空间

$$\mathcal{F}_s := \bigcup_{\substack{S \subset [p], \\ |S| \leq s}} \mathcal{H}_S \tag{4}$$

对于每个 $1 \leq s \leq p$。

接下来，[17, 25], 将（1）从函数空间转换到序列空间是很方便的。张量积 $L^2([0,1])^{\otimes p}$ 具有标准正交基 $\{\Phi_\ell\}_{\ell \in \mathbb{N}^p}$，其中对于$t \in [0,1]^p$，$\Phi_\ell(t) = \prod_{j=1}^p \psi_{\ell_j}(t_j)$。为简化记号，记 $\phi_{j,\,k}(t) = \psi_k(t_j)$，对于 $k \in \mathbb{N}$ 和 $j \in [p]$。定义随机变量

$$X_{k,j} := \int_{[0,1]^p} \phi_{j,k}(x)\, dY_x \sim N(\theta_{k,j}, n^{-1}) \tag{5}$$

其中 $\theta_{k,\,j} = \int_0^1 \psi_k(x) f_j(x)\, dx$。这里使用了对所有 j 均成立的假设 $\int_0^1 f_j(t)\, dt = 0$。注意，由正交性可知，$\{X_{k,j}\}_{k \in \mathbb{N},\ j \in [p]}$ 是独立随机变量的集合。记号$\Theta \in \mathbb{R}^{\mathbb{N} \times p}$ 将经常用于表示完整的系数矩阵。记号 $\Theta_j \in \mathbb{R}^{\mathbb{N}}$ 也将用于表示 Θ的第 j列。对于 $f \in \mathcal{F}_s$，相应的系数集合为

$$\mathcal{T}_s := \left\{ \Theta \in \mathbb{R}^{\mathbb{N} \times p} : \sum_{j=1}^p \mathbb{I}_{\{\Theta_j \neq 0\}} \leq s \text{ and } \sum_{k=1}^\infty \frac{\theta_{k,j}^2}{\mu_k} \leq 1 \text{ for all } j \right\}. \tag{6}$$

注意，参数空间 $\mathcal{F}_s$ 和 $\mathcal{T}_s$ 是相互对应的；在同一语境中，我们常常不加说明地交替使用 f 和 Θ。约定 f 是一个函数，而 Θ 是其对应的基系数。记号 E_f、P_f、E_Θ 和 P_Θ 用于表示相对于所指参数的期望和概率测度。

1.2.3 问题

如前所述，给定来自（1）的一个观测，信号检测问题（2）-（3）值得研究。目标是刻画非渐近极小极大分离率 $\varepsilon^*_{\mathcal{H}} = \varepsilon^*_{\mathcal{H}}(p,\ s,\ n)$。

定义 1. *We say $\varepsilon^*_{\mathcal{H}} = \varepsilon^*_{\mathcal{H}}(p,s,n)$ is the nonasymptotic minimax separation rate for the problem (2)-(3) if*

(i) for all $\eta \in (0,1)$ there exists $C_\eta > 0$ depending only on η such that for all $C > C_\eta$,

$$\inf_\varphi \left\{ P_0\{\varphi \neq 0\} + \sup_{\substack{f \in \mathcal{F}_s, \\ ||f||_2 \geq C\varepsilon^*_{\mathcal{H}}(p,s,n)}} P_f\{\varphi \neq 1\} \right\} \leq \eta,$$

(ii) for all $\eta \in (0,1)$ *there exists* $c_\eta > 0$ *depending only on* η *such that for all* $0 < c < c_\eta$,

$$\inf_{\varphi} \left\{ P_0\{\varphi \neq 0\} + \sup_{\substack{f \in \mathcal{F}_s, \\ ||f||_2 \geq c\varepsilon^*_{\mathcal{H}}(p,s,n)}} P_f\{\varphi \neq 1\} \right\} \geq 1 - \eta.$$

非渐近极小极大检验框架由 Baraud 开创 [2]，它刻画了检测问题的基本统计极限（精确到通用因子）。该框架之所以是非渐近的，是因为极小极大分离率的条件对问题参数的任意配置均成立。

由于 $||f||_2^2 = \sum_{j=1}^{p} \sum_{k=1}^{\infty} \theta_{k,\,j}^2$，该问题可以等价地表述为序列空间中的形式

$$H_0 : \Theta \equiv 0, \tag{7}$$

$$H_1 : ||\Theta||_F \geq \varepsilon \text{ and } \Theta \in \mathscr{T}_s. \tag{8}$$

这个检验问题及其参数空间（6），即使脱离稀疏加性模型和再生核希尔伯特空间的背景，本身也很有趣。事实上，检验问题（7）-（8）本质上是 Baraud 的开创性论文 [2] 中所考虑的椭球上的检测问题的组稀疏扩展。实际上，这种解读正是我们最初研究该检测问题的动机。与稀疏加性模型的联系则是后来才考虑的；这类似于后来的论文[17] 在早期基础性工作 [25] 之上，研究稀疏加性模型的做法；该早期工作研究的是稀疏多通道（本质上是组稀疏）模型。以序列问题的视角来研究非参数回归由来已久 [2, 5, 15, 24, 25, 27, 40, 48]，这不仅源于其深刻的内在联系，也因为这种视角能将问题提炼至本质，并排除技术性干扰。我们的结果完全可以放在序列问题（7）-（8）的背景下阅读（我们也鼓励读者这样做）。

2 极小极大速率

我们首先介绍一些高层次且直观的想法，然后在第 2.1 节非正式地陈述主要结果。第 2.2 节将推导若干关键量。在第 2.3 节和 2.4 节中，我们将分别正式给出极小极大下界和上界。第 2.5 节将讨论一些说明一般结果的特殊情形。

2.1 一个朴素的设想

首先想到的是参考以往的结果，从中获取背景信息，并尝试猜测最优检验速率。为说明这种思路可能如何展开，考虑一维单位区间上的经典高斯白噪声模型：

$$dY_x = f(x)\,dx + \frac{1}{\sqrt{n}} dB_x$$

对于 $x \in [0,1]$。假设 f 位于再生核希尔伯特空间 $\mathcal{H}$ 的单位球内，因此可在相应的标准正交基下进行分解，其系数向量 $\theta \in \ell^2(\mathbb{N})$ 位于由核的特征值 $\mu_1 \geq \mu_2 \geq \ldots \geq 0$ 确定的椭球内。最优估计速率由 Birg´e 和 Massart [4] 给出

$$\epsilon_{\text{est}}^2 \asymp \max_{\nu \in \mathbb{N}} \left\{ \mu_\nu \wedge \frac{\nu}{n} \right\}.$$

Baraud [2] 确立了信号检测问题的极小极大分离率

$$H_0 : f \equiv 0,$$
$$H_1 : ||f||_2 \geq \varepsilon \text{ and } ||f||_{\mathcal{H}} \leq 1$$

表示为

$$\epsilon_{\text{test}}^2 \asymp \max_{\nu \in \mathbb{N}} \left\{ \mu_\nu \wedge \frac{\sqrt{\nu}}{n} \right\}.$$

在估计和检验中，最大化元 ν^* 都可以被理解为正确截断阶数。具体而言，对于检验，Baraud方法 [2], 在序列空间中运行，当$\sum_{k=1}^{\nu^*} X_k^2 - \frac{\nu^*}{n} \gtrsim \frac{\sqrt{\nu^*}}{n}$ 时拒绝 H_0 。不使用 $k > \nu^*$ 的数据，正是在这个意义上， ν^* 被理解为截断水平。为说明这些结果，具有光滑性 α 的Sobolev空间的速率为 $\epsilon_{\text{est}}^2 \asymp n^{-\frac{2\alpha}{2\alpha+1}}$ 和 $\epsilon_{\text{test}}^2 \asymp n^{-\frac{4\alpha}{4\alpha+1}}$ ，因为 $\mu_\nu \asymp \nu^{-2\alpha}$。如今， [2, 4]的这些非渐近结果已广为人知，并成为经典结果。

转向模型（1）中的稀疏加性信号设定，Raskutti 等人 [38] 推导出了估计的非渐近极小极大速率

$$\frac{s \log p}{n} + s\epsilon_{\text{est}}^2.$$

虽然他们的上界适用于任意选择的 $\mathcal{H}$ （满足一些温和条件）和稀疏度 s，但只有当 $s = o(p)$，且 $\mathcal{H}$ 的单位球具有对数或多项式标度度量熵时，才能得到匹配的下界。Raskutti 等人[38] 得到的这一收敛速率令人满意。很容易直观地看出，由于参数空间中的稀疏性，会出现 ${}^{s}\log_{n}{}^{p}$ 这一项。项 $s\epsilon_{\text{est}}^2$ 是估计 $\mathcal{H}$ 中的 s 个单变量函数时的自然收敛速率，仿佛稀疏模式已知一样。值得注意的是， $\mathcal{H}$ 的选择与稀疏结构之间没有相互作用。稀疏项 ${}^{s}\log_{n}{}^{p}$ 与 $\mathcal{H}$ 独立，而估计项 $s\epsilon_{\text{est}}^2$ 与维数无关。

人们可能会直觉地认为，这种缺乏交互作用的情况普遍存在。直觉或许会认为，稀疏加性模型中的信号检测也不过是 s 个单变量非参数检测问题，再加上检测一个在未知稀疏模式上非零的信号这一问题。若将其视为两个不同的问题，人们可能会猜想，最优检验速率应为$s\epsilon_{\text{test}}^2$ 加上 s-稀疏检测率。Collier等人 [10] 给出了 s-稀疏检验速率的一个自然候选值，即速率为$s\log\left(\frac{p}{s^2}\right)$，适用于 $s < \sqrt{p}$；而 $\frac{\sqrt{p}}{n}$，适用于 $s \geq \sqrt{p}$。

然而，只需快速查看 [25] ，即可发现该猜想在索об列夫 $\mathcal{H}$ 的情形下并不成立。尽管 [25] 很快就打消了乐观的想法，却几乎没有揭示稀疏性与 $\mathcal{H}$ 之间的相互作用。我们的结果明确刻画了这种非平凡的相互作用，适用于一般的 $\mathcal{H}$。我们证明，极小极大分离率为

$$\varepsilon_{\mathcal{H}}^*(p,s,n)^2 \asymp s \wedge \begin{cases} \frac{s\log\left(\frac{p}{s^2}\right)}{n} + s \cdot \max_{\nu\in\mathbb{N}} \left\{ \mu_\nu \wedge \frac{\sqrt{\nu \log\left(1+\frac{p}{s^2}\right)}}{n} \right\} & \text{if } s < \sqrt{p}, \\ s \cdot \max_{\nu\in\mathbb{N}} \left\{ \mu_\nu \wedge \frac{\sqrt{\nu \log\left(1+\frac{p}{s^2}\right)}}{n} \right\} & \text{if } s \geq \sqrt{p}. \end{cases} \tag{9}$$

该速率与 Collier等人 [10] 的稀疏检验速率以及 Baraud [2], 的非参数检验速率有些相似，但将二者结合起来并非 *a priori* 那么简单。在讨论的这一阶段，我们尚未充分展开，因而无法解释该速率形式的由来。本文后文将进一步讨论该速率的各种特征。

2.2 预备知识

本节将定义一些关键部分。

定义2。 *Define* $\Gamma_{\mathcal{H}} = \Gamma_{\mathcal{H}}(p,s,n)$ *to be the quantity*

$$\Gamma_{\mathcal{H}} := \max_{\nu\in\mathbb{N}} \left\{ \mu_\nu \wedge \frac{\sqrt{\nu \log\left(1+\frac{p}{s^2}\right)}}{n} \right\}.$$

注意，由于 $\mu_1 = 1$，可得 $\Gamma_{\mathcal{H}} \gtrsim \frac{1}{n}$。从 (9) 可以很容易看出，需要考虑两种宽泛的情形。当 $n \lesssim \log(1+p/s^2)$ 时，我们有 $\varepsilon^*_{\mathcal{H}}(p,s,n)^2 \asymp s$。在 $n \gtrsim \log(1+p/s^2)$ 这一情形中，速率则更为复杂。第一种情形实际上是平凡的，因为任何信号 $f \in \mathcal{F}_s$ 都必须满足 $||f||_2^2 \le s$，这是由 $\mu_1 = 1$ 所致。因此，始终接受 H_0 的退化检验可以平凡地检测出 $||f||_2^2 \ge 2s$ 的稀疏加性信号。因此，上界 $\varepsilon^*_{\mathcal{H}}(p,s,n)^2 \lesssim s$ 可以平凡地达到。事实证明，可以证明一个匹配的下界，从而确立 $n \lesssim \log(1+p/s^2)$ 这一情形本质上是平凡的；正式表述见第 2.3 节。

更一般地，在讨论下界时，形式（9）很有用。讨论上界时，另一种形式更为方便；这种形式在 [10, 11]的语境中很常见。

定义3. *Define* $\nu_{\mathcal{H}}$ *to be the smallest positive integer* ν *such that*

$$\mu_\nu \le \frac{\sqrt{\nu \log\left(1+\frac{p}{s^2}\right)}}{n}.$$

正如下一个引理所示， $\nu_{\mathcal{H}}$ 本质上是定义 $\Gamma_{\mathcal{H}}$中最大化问题的解。类比 Baraud [2] 在第2.1节 $\nu_{\mathcal{H}}$ 所述的结果，可以将其理解为依据维数和稀疏性确定的恰当截断阶。

引理 1. *If* 对数$\left(1+\frac{p}{s^2}\right) \le \frac{n}{2}$， *then* $\Gamma_{\mathcal{H}} \le \frac{\sqrt{\nu_{\mathcal{H}} \frac{\text{对数}\left(1+\frac{p}{s^2}\right)}{n}}}{n} \le \sqrt{2}\Gamma_{\mathcal{H}}$.

根据引理 1，检验速率可以表示为

$$\varepsilon^*_{\mathcal{H}}(p,s,n)^2 \asymp \begin{cases} \frac{s}{n}\log\left(1+\frac{p}{s^2}\right) + \frac{s}{n}\sqrt{\nu_{\mathcal{H}} \log\left(1+\frac{p}{s^2}\right)} & \text{if } s < \sqrt{p}, \\ \frac{\sqrt{p\nu_{\mathcal{H}}}}{n} & \text{if } s \ge \sqrt{p}. \end{cases} \tag{10}$$

引理 $\left(1+\frac{p}{s^2}\right) \lesssim n$ 中的对数条件 1 是自然的，因为当$n \lesssim$ 对数$\left(1+\frac{p}{s^2}\right)$时会出现平凡性。

2.3 下界

本节将正式陈述极小极大分离率的一个下界。定义 ne

$$\psi(p,s,n)^2 := \begin{cases} \frac{s\log\left(1+\frac{p}{s^2}\right)}{n} \vee s\Gamma_{\mathcal{H}} & \text{if } s < \sqrt{p}, \\ s\Gamma_{\mathcal{H}} & \text{if } s \ge \sqrt{p}. \end{cases} \tag{11}$$

首先证明，当 $\log\left(\frac{p}{s^2}\right) \gtrsim n$ 时，检验问题是平凡的。

命题 1（平凡性）。 *Suppose* $1 \le s \le p$. *Suppose* $\kappa > 0$ *and* $\log\left(1+\frac{p}{s^2}\right) \ge \kappa n$. *If* $\eta \in (0,1)$, *then for any* $0 < c < 1 \wedge \sqrt{\kappa} \wedge \sqrt{\kappa \log(1+4\eta^2)}$ *we have*

$$\inf_{\varphi} \left\{ P_0\{\varphi \ne 0\} + \sup_{\substack{f\in\mathcal{F}_s, \\ ||f||_2 \ge c\sqrt{s}}} P_f\{\varphi \ne 1\} \right\} \ge 1-\eta.$$

命题 1 断言，为了实现较小的检验误差，一个必要条件是$||f||_2 \ge C\sqrt{s}$ 对于充分大的 $C > 0$。要理解命题 1 为何是关于平凡性的陈述，请注意 $\mathcal{F}_s \subset \left\{f \in L^2([0,1]^p) : ||f||_2 \le \sqrt{s}\right\}$。换言之，模型（ $f \in \mathcal{F}_s$ （1）中的所有候选 $\sqrt{s}$都位于半径为 $f \in \mathcal{F}_s$ 的球内。当 $\log\left(\frac{p}{s^2}\right) \gtrsim n$时，基本不存在具有可检测范数的函数，因此该问题本质上是平凡的。

下界构造十分直接。在序列空间中，考虑一个带有先验 π 的备择假设，其中一次抽样 $\Theta \sim \pi$ 的获得方式为：从所有集合中均匀随机地抽取一个大小为 s 的子集 $S \subset [p]$，并在 $k = 1$ 时令 $\theta_{k,\ j} = \rho$，否则令 $\theta_{k,\ j} = 0$。ρ 的取值决定了零假设与备择假设之间的分离度，因为 $||\Theta||_F^2 = s\rho^2$。然而，为了确保 π 的支撑集包含于参数空间，且零假设与备择假设之间不可能实现渐近零错误的区分，ρ 必须满足某些约束。注意，这一构造使我们处于一维稀疏序列模型中（因为对于 $k \geq 2$，有 $\theta_{k,\ j} = 0$），而这恰好是 [10] 所研究的情形。根据他们的结果可知，我们必须满足 $\rho^2 \lesssim \log\left(1 + \frac{p}{s^2}\right)/n$。为了确保 π 的支撑集包含于参数空间，对于所有 $j \in [p]$，我们必须满足 $\sum_{k=1}^{\infty} \theta_{k,\ j}^2/\mu_k \leq 1$。由于 $\mu_1 = 1$，可得 $\sum_{k=1}^{\infty} \theta_{k,\ j}^2/\mu_k = \theta_{1,\ j}^2 \leq \rho^2$，因此必须施加约束 $\rho \lesssim 1$。当 $\log(1 + p/s^2) \gtrsim n$ 时，只有第二个条件 $\rho \lesssim 1$ 起约束作用，因此我们所能达到的最大分离度为 $||\Theta||_F^2 \asymp s$。因此，在这一情形下，该问题是平凡的。

为确保非平凡性，假设 $\log\left(1 + \frac{p}{s^2}\right) \leq \frac{n}{2}$。选择因子 $1/2$ 仅为方便起见，并非必要。事实上，条件 $\log(1 + \frac{p}{s^2}) < n$ 始终足以满足我们的目的，而条件 $\log\left(1 + \frac{p}{s^2}\right) \leq n$ 对于 $n > 1$ 也足够。下述定理给出了 $\varepsilon_{\mathcal{H}}^*(p, s, n)^2 \gtrsim \psi(p, s, n)^2$。

定理 1. *Suppose* $1 \leq s \leq p$ *and* $\log\left(1 + \frac{p}{s^2}\right) \leq \frac{n}{2}$. *If* $\eta \in (0,1)$ *then there exists* $c_\eta > 0$ *depending only on* η *such that for any* $0 < c < c_\eta$ *we have*

$$\inf_{\varphi} \left\{ P_0\{\varphi \neq 0\} + \sup_{\substack{f \in \mathcal{F}_s, \\ ||f||_2 \geq c\psi(p,s,n)}} P_f\{\varphi \neq 1\} \right\} \geq 1 - \eta$$

where ψ *is given by (11).*

下界通过勒卡姆二点方法（也称为“模糊”假设方法[45]）得到证明，这是极小极大假设检验文献中的标准方法。[24]论证中采用的先验分布是一种自然构造。具体而言，活跃集（即非零坐标函数的大小 s集合）通过均匀随机方式选取，相应的非零坐标函数则从文献中常用的一元非参数先验中抽取。[2, 24]。

2.4 上界

本节将构造检验程序，并给出确立速率最优性的形式化陈述。式（10）中的速率形式更便于使用；阅读上界相关陈述时应牢记这一点。

2.4.1 稀疏区域中的硬阈值化

本节讨论稀疏区域 $s < \sqrt{p}$。对于任意 $d \in \mathbb{N}$ 和 $j \in [p]$，定义 $E_j(d) = n\sum_{k \leq d} X_{k,\ j}^2$ 其中，数据 $\{X_{k,j}\}_{k \in \mathbb{N},\ j \in [p]}$ 通过变换到序列空间（5）定义。对于任意 $r \geq 0$，定义

$$T_r(d) := \sum_{j=1}^{p} (E_j(d) - \alpha_r(d))\, \mathbb{I}_{\{E_j(d) \geq d + r^2\}} \tag{12}$$

其中

$$\alpha_r(d) := \frac{E\left(||g||^2\, \mathbb{I}_{\{||g||^2 \geq d + r^2\}}\right)}{P\{||g||^2 \geq d + r^2\}} \tag{13}$$

其中 $g \sim N(0,\ I_d)$。注意 $\alpha_r(d)$ 是原假设下的条件期望。随机变量 $T_r(d)$ 将用作检验统计量。Collier等人最早定义了这种统计量 [10]，类似的统计量也已成功应用于其他信号检测问题 [8, 9, 30,34]。然而，文献中以往的所有研究都只在 $d = 1$下使用这一统计量。对于我们的问题，则必须让 d不断增大。因此，需要进行更精细的分析，而且正如我们稍后将讨论的，上界中会出现本质上全新的现象。如第2.2节所述，

量 $\nu_{\mathcal{H}}$ 可以理解为正确截断阶数，也就是说，事实证明正确的选择是 $d \asymp \nu_{\mathcal{H}}$。 r 的选择更加复杂，实际上，根据 d 的大小存在两个不同的情形。当 $d \gtrsim \log\left(1+\frac{p}{s^2}\right)$ 时，这种情形称为“主体部分”情形。与之互补的情形 $d \lesssim \log\left(1+\frac{p}{s^2}\right)$ 则称为“尾部”情形。

命题 2 (主体部分)。 *Set $d = \nu_{\mathcal{H}} \vee \lceil D \rceil$ where D is the universal constant from Lemma 15. There exist universal positive constants K_1, K_2, and K_3 such that the following holds. Suppose $1 \le s < \sqrt{p}$ and 对数 $\left(1+\frac{p}{s^2}\right) \le K_3^2 d$. Set*

$$\tau(p,s,n)^2 = \frac{s\sqrt{\nu_{\mathcal{H}} \log\left(1+\frac{p}{s^2}\right)}}{n}.$$

If $\eta \in (0,1)$ then there exists $C_\eta > 0$ depending only on η such that for all $C > C_\eta$, we have

$$P_0\left\{T_r(d) > CK_1 n\tau(p,s,n)^2\right\} + \sup_{\substack{f\in\mathcal{F}_s,\\ \|f\|_2 \ge C\tau(p,s,n)}} P_f\left\{T_r(d) \le CK_1 n\tau(p,s,n)^2\right\} \le \eta$$

where $r = K_2\left(d \text{对数}\left(1+\frac{p}{s^2}\right)\right)^{1/4}$. Here, $T_r(d)$ is given by (12).

命题 3（尾部）。 *Set $d = \nu_{\mathcal{H}} \vee \lceil D \rceil$ where D is the universal constant from Lemma 11. Let K_3 denote the universal positive constant from Proposition 2. There exist universal positive constants K_1 and K_2 such that the following holds. Suppose $1 \le s < \sqrt{p}$ and 对数 $\left(1+\frac{p}{s^2}\right) \ge K_3^2 d$. Set*

$$\tau(p,s,n)^2 = \frac{s\log\left(1+\frac{p}{s^2}\right)}{n}.$$

If $\eta \in (0,1)$ then there exists $C_\eta > 0$ depending only on η such that for all $C > C_\eta$, we have

$$P_0\left\{T_r(d) > CK_1 n\tau(p,s,n)^2\right\} + \sup_{\substack{f\in\mathcal{F}_s,\\ \|f\|_2 \ge C\tau(p,s,n)}} P_f\left\{T_r(d) \le CK_1 n\tau(p,s,n)^2\right\} \le \eta$$

where $r = K_2\sqrt{\text{对数}\left(1+\frac{p}{s^2}\right)}$. Here, $T_r(d)$ is given by (12).

命题 3 和 2 由此可得，对于 $s < \sqrt{p}$，极小极大分离率的上界为 s对数 $\left(1+\frac{p}{s^2}\right)/n + s\sqrt{\nu_{\mathcal{H}}}$对数 $\left(1+\frac{p}{s^2}\right)/n$。由引理 1，可知

$$\varepsilon^*_{\mathcal{H}}(p,s,n) \lesssim \frac{s\log\left(1+\frac{p}{s^2}\right)}{n} + s\Gamma_{\mathcal{H}}$$

在满足条件 $\log\left(1+\frac{p}{s^2}\right) \le \frac{n}{2}$ 的情况下。正如第 1 节 2.3中的命题所证明的那样，这样的条件对于避免平凡性至关重要。

命题 3 和 2 揭示了极小极大分离率在该点处发生了一个有趣的相变

$$\nu_{\mathcal{H}} \asymp \log\left(1+\frac{p}{s^2}\right).$$

这种相变现象由 χ^2_d 的尾部行为驱动。考虑在零假设分布下 $E_j(d) \sim \chi^2_d$ 对所有 j均成立，因此统计量 $T_r(d)$ 是 p 独立且经过阈值化的 χ^2_d 随机变量之和。根据 Laurent 和 Massart 的一个著名引理 [31]（利用 Bernstein 不等式也可在常数意义下得到），对于任意 $u > 0$ 都有

$$P\left\{\chi^2_d - d \ge 2\sqrt{du} + 2u\right\} \le e^{-u}. \tag{14}$$

粗略地说，当处于“主体部分”时，$\chi^2_d - d$ 表现出次高斯型偏差 $u \lesssim \sqrt{d}$；而在“尾部” $u \gtrsim \sqrt{d}$表现出次指数型偏差。因此，当 $r \lesssim \sqrt{d}$时，应将 $T_r(d)$理解为阈值化次高斯随机变量之和；当$r \gtrsim \sqrt{d}$时，则应理解为阈值化次指数随机变量之和。

考察（14），在“尾部” $u \gtrsim \sqrt{d}$ 可得 $2\sqrt{du} + 2u \asymp u$，它不再依赖于 d。类似地，在命题3 中，阈值取为 $r \gtrsim \sqrt{d}$，因此所得速率不依赖于 d，进而也不依赖于 $\mathcal{H}$。另一方面，在

“主体部分” $u \lesssim \sqrt{d}$ 可得 $2\sqrt{du}+2u \asymp \sqrt{du}$。类似地，在命题2 中，阈值取为$r \lesssim \sqrt{d}$，因此所得速率确实依赖于 d，从而依赖于 $\mathcal{H}$。

2.4.2 χ^2 稠密情形下的检验

在稠密区域，情况较为简单$s \geq \sqrt{p}$，因为只需使用 χ^2 检验统计量 $T := n\sum_{j=1}^{p}\sum_{k\leq \nu_{\mathcal{H}}} X_{k,j}^2$。

命题 4。 *Suppose $s \geq \sqrt{p}$. Set*

$$\tau(p,s,n)^2 = \frac{\sqrt{p\nu_{\mathcal{H}}}}{n}.$$

If $\eta \in (0,1)$ then there exists $C_\eta > 0$ depending only on η such that for all $C > C_\eta$, we have

$$P_0\left\{T > p\nu_{\mathcal{H}} + \frac{2}{\sqrt{\eta}}\sqrt{p\nu_{\mathcal{H}}}\right\} + \sup_{\substack{f\in\mathcal{F}_s,\\ ||f||_2 \geq C\tau(p,s,n)}} P_f\left\{T \leq p\nu_{\mathcal{H}} + \frac{2}{\sqrt{\eta}}\sqrt{p\nu_{\mathcal{H}}}\right\} \leq \eta.$$

与稀疏情形一样，(10) 引理 1 断言 $\tau(p,s,n)^2 \asymp s\Gamma_{\mathcal{H}}$ 只要满足条件 $\log\left(1+\frac{p}{s^2}\right) \leq n/2$ 即可。

2.5 特殊情形

在正式给出极小极大分离率的下界和上界后，探讨若干特殊情形并观察各种现象颇有意义。在以下示例中，均假设 $\log\left(1+\frac{p}{s^2}\right) \leq n/2$，如前文所述。

2.5.1 索ob列夫

以 $\mu_k \asymp k^{-2\alpha}$ 为代表来说明光滑性为 $\alpha > 0$ 的 Sobolev空间情形，可得极小极大分离率

$$\varepsilon_{\mathcal{H}}^*(p,s,n)^2 \asymp \begin{cases} \frac{s\log\left(\frac{p}{s^2}\right)}{n} + s\left(\frac{n}{\sqrt{\log\left(\frac{p}{s^2}\right)}}\right)^{-\frac{4\alpha}{4\alpha+1}} & \text{if } s < \sqrt{p}, \\ s\left(\frac{ns}{\sqrt{p}}\right)^{-\frac{4\alpha}{4\alpha+1}} & \text{if } s \geq \sqrt{p}. \end{cases}$$

将所得分离率与 Ingster 和 Lepski 的结果进行比较很有意义（[25]中的定理 2）（另见 [17]），不过，他们对参数空间的选择与本文考虑的略有不同。在稀疏情形 $s = O\left(p^{1/2-\delta}\right)$ 下，对于常数 $\delta \in (0, 1/2]$，他们的分离率为

$$\begin{cases} s\left(\frac{n}{\sqrt{\log p}}\right)^{-\frac{4\alpha}{4\alpha+1}} & \text{if } \log p \leq n^{\frac{1}{2\alpha+1}}, \\ s\frac{\log p}{n} & \text{if } \log p > n^{\frac{1}{2\alpha+1}}. \end{cases}$$

在稠密区域 $p = O(s^2)$ 中，他们的分离率（ [25]中的定理 1）为 $s\left(\frac{ns}{\sqrt{p}}\right)^{-\frac{4\alpha}{4\alpha+1}}$。简单的代数运算可验证，在这些稀疏性情形下，我们的分离率确实与 Ingster 和 Lepski 的分离率一致。

[1]在稀疏区域中，其速率在 $\log p \asymp n^{\frac{1}{2\alpha+1}}$ 处出现的奇特相变，依据我们的结果现在有了合理的解释。情形 $\log p \lesssim n^{\frac{1}{2\alpha+1}}$ 对应于“主体部分”情形，此时 $T_r(d)$ 中的 (12) 表现得像阈值化次高斯随机变量之和。在另一侧，即 $\log p \gtrsim n^{\frac{1}{2\alpha+1}}$ 时，$T_r(d)$ 呈现次指数行为。事实上，我们的结果在 $s \lesssim p^{1/2-\delta}$ 之外给出了更多细节。只需假设 $s < \sqrt{p}$ （例如，现在允许 $s = \frac{\sqrt{p}}{\log p}$ ）。那么，“ 主体部分 ” 与“尾部”情形之间的相变实际上发生在 $\log\left(1+\frac{p}{s^2}\right) \asymp n^{\frac{1}{2\alpha+1}}$ 处。

2.5.2 有限维

考虑有限维情形，即 $1=\mu_1=\mu_2=...=\mu_m>\mu_{m+1}=\mu_{m+2}=...=0$ 其中某个正整数 m。具有此类结构的函数空间包括线性函数、有限次多项式，以及通常基于有限秩核函数的再生核希尔伯特空间。如果 $m<$ $n^2\log\left(1+\frac{p}{s^2}\right)$，极小极大分离率为

$$\varepsilon_{\mathcal{H}}^{*}(p,s,n)^2 \asymp \begin{cases} \frac{s\log\left(\frac{p}{s^2}\right)}{n}+\frac{s\sqrt{m\log\left(\frac{p}{s^2}\right)}}{n} & if\ s<\sqrt{p}, \\ \frac{\sqrt{pm}}{n} & if\ s\geq\sqrt{p}. \end{cases}$$

在稀疏区域中，主体部分与尾部区域之间的相变发生在 $\log\left(1+\frac{p}{s^2}\right)\asymp m$。

2.5.3 指数衰减

作为另一个例子，考虑特征值 $\mu_k=c_1e^{-c_2k^\gamma}$ 的指数衰减，其中 c_1、c_2 是通用常数，且 $\gamma>0$。这种衰减是基于高斯核的再生核希尔伯特空间的一个特征。极小极大分离率为

$$\varepsilon_{\mathcal{H}}^{*}(p,s,n)^2 \asymp \begin{cases} \frac{s\log\left(\frac{p}{s^2}\right)}{n}+s\log^{\frac{1}{2\gamma}}\left(\frac{n}{\sqrt{\log\left(\frac{p}{s^2}\right)}}\right)\cdot\frac{\sqrt{\log\left(\frac{p}{s^2}\right)}}{n} & if\ s<\sqrt{p} \\ \frac{\sqrt{p}}{n}\log^{\frac{1}{2\gamma}}\left(\frac{ns}{\sqrt{p}}\right) & if\ s\geq\sqrt{p}. \end{cases}$$

在稀疏区域中，主体部分与尾部区域之间的相变发生在 $\log^{\gamma}\left(1+\frac{p}{s^2}\right)\asymp\log n$。极小极大分离率非常接近有限维情形下的分离率。鉴于基于高斯核的再生核希尔伯特空间众所周知属于相当“小”的非参数函数空间，这一点是合理的 [47]。

3 自适应

到目前为止，稀疏参数 s 一直被假定为已知，所构造的检验也利用了这一信息。在实践中，统计学家通常并不知道稀疏度，因此，研究是否能够构造自适应检验便很有意义。在本节中，我们将建立一个适用于一般 $\mathcal{H}$的自适应检验率；事实证明，自适应代价会微妙地取决于函数空间。

据我们所知，Spokoiny的文章首次证明了信号检测问题中不可避免的自适应代价 [41]。后续研究表明，在各种情形下进行自适应检验都存在不可避免的代价（参见 [24] 及其中的参考文献）。早期研究主要关注在单变量非参数回归框架下对未知平滑参数进行自适应。近年来，人们开始研究高维模型中的对稀疏性的自适应。在许多问题中，可以在不付出速率代价的情况下实现对稀疏性的自适应（对于某些问题，甚至无需付出常数代价）[1, 14, 18, 26, 30, 34]。在Sobolev空间中的稀疏加性模型背景下，Ingster和Lepski[25] （另见 [17]）研究了自适应问题，我们将在第 3.4节讨论他们的结果。

3.1 预备知识

为了正式陈述我们的结果，需要对第 2.2 节中的概念作一些细微推广。

定义4. *For $a>0$, define $\nu_{\mathcal{H}}(s,a)$ to be the smallest positive integer ν satisfying*

$$\mu_\nu\leq\frac{\sqrt{\nu\log\left(1+\frac{pa}{s^2}\right)}}{n}.$$

定义 5. *For* $a > 0$, *define* $\Gamma_{\mathcal{H}}(s,a)$ *to be*

$$\Gamma_{\mathcal{H}}(s,a) := \max_{\nu \in \mathbb{N}} \left\{ \mu_\nu \wedge \frac{\sqrt{\nu \log\left(1 + \frac{pa}{s^2}\right)}}{n} \right\}.$$

注意 $\nu_{\mathcal{H}}(s, a)$ 随 a 增大而递减，且 $\Gamma_{\mathcal{H}}(s, a)$ 随 a增大而递增。如第 2.2 节所讨论，在讨论分离率时，两种不同形式都很有用，而引理 1 为这种讨论提供了便利。下面的引理是对其的轻微推广，目的相同。

引理 2. *Suppose* $a > 0$. *If* $\log\left(1 + \frac{pa}{s^2}\right) \le \frac{n}{2}$, *then*

$$\Gamma_{\mathcal{H}}(s,a) \le \frac{\sqrt{\nu_{\mathcal{H}}(s,a) \log\left(1 + \frac{pa}{s^2}\right)}}{n} \le \sqrt{2}\Gamma_{\mathcal{H}}(s,a).$$

从较高层面看，自适应的核心问题在于选择估计量或检验的超参数（例如带宽或惩罚项）。通常存在某个最优选择，但要选定它，需要了解某个未知参数（如光滑性或稀疏性）。在当前问题中，由于稀疏度 s 未知， ν 的最优选择也未知。

在自适应检验中，典型策略是固定一组不同取值的网格，为超参数的每个可能取值构造一个检验，并在检测时使用这组检验中的最大值对应的检验。通常会选择一个几何网格，而网格基数的对数反映了检验程序所付出的代价 [25, 34, 41]。事实证明，这一高层次直觉同样适用于稀疏加性模型中的信号检测，但由于必须适应一个通用的 $\mathcal{H}$ ，如何选择网格的细节并不直接。对于 $a > 0$，定义

$$\mathscr{V}_a := \left\{ 2^k : k \in \mathbb{N} \cup \{0\} \text{ and } 2^{k-1} < \nu_{\mathcal{H}}(s,a) \le 2^k \text{ for some } s \in [p] \right\}.$$

定义

$$\mathscr{A}_{\mathcal{H}} := \sup\{a \ge 1 : \log(e|\mathscr{V}_a|) \ge a\} \tag{15}$$

并定义

$$\mathscr{V}_{\mathcal{H}} := \mathscr{V}_{\mathscr{A}_{\mathcal{H}}}. \tag{16}$$

将使用网格 $\mathscr{V}_{\mathcal{H}}$ 。显然， $\mathscr{A}_{\mathcal{H}}$ 是有限的，因为粗略上界 $\mathscr{A}_{\mathcal{H}} \le \log(ep)$ 立即可得。下面的引理表明， $\mathscr{A}_{\mathcal{H}}$ 本质上是一个不动点。

引理 3. *If* $\log(1 + p\mathscr{A}_{\mathcal{H}}) \le \frac{n}{2}$, *then* $\mathscr{A}_{\mathcal{H}} \le \log(e|\mathscr{V}_{\mathcal{H}}|) \le 2\mathscr{A}_{\mathcal{H}}$.

结果表明，可以构造一个自适应检验程序，其代价由 $\mathscr{A}_{\mathcal{H}}$ 决定（等价地，在常数因子范围内由 $\log(e|\mathscr{V}_{\mathcal{H}}|)$ 决定）。具体而言，假设 $\log(1 + p\mathscr{A}_{\mathcal{H}}) \le n/2$。我们将构造一个自适应检验，使其达到

$$\begin{cases} \frac{s}{n} \log\left(1 + \frac{p\mathscr{A}_{\mathcal{H}}}{s^2}\right) + s\Gamma_{\mathcal{H}}(s, \mathscr{A}_{\mathcal{H}}) & \text{if } s < \sqrt{p\mathscr{A}_{\mathcal{H}}}, \\ s\Gamma_{\mathcal{H}}(s, \mathscr{A}_{\mathcal{H}}) & \text{if } s \ge \sqrt{p\mathscr{A}_{\mathcal{H}}}. \end{cases}$$

这种分离率的形式与极小极大速率非常相似，只是相变移到了 $s \asymp \sqrt{p\mathscr{A}_{\mathcal{H}}}$ ，并且需要付出涉及 $\mathscr{A}_{\mathcal{H}}$ 的代价。 $\mathscr{A}_{\mathcal{H}}$ 的取值会随着 $\mathcal{H}$的选择而有很大变化，下面第 3.4节将举例说明。

我们选择几何网格来得到引理 3 的不动点刻画，这一点可能乍看之下并不直观。从整体上看，可以通过注意到其中存在两个相互竞争的因素来理解这一选择。首先，固定 $a \ge 1$ ，并注意到它可以被理解为表示扫描某个网格所需的目标成本。为了达到这一目标成本，我们应对每个稀疏性 s使用截断水平 $\nu_{\mathcal{H}}(s, a)$。现在，考虑我们将扫描由这些截断水平得到的几何网格 $\mathscr{V}_a$ 。如果该几何网格的基数所对应的成本低于 a，那么，为了达到与更大网格相关联的目标成本，却扫描一个更小的网格，似乎就很奇怪了。直观上，我们或许会尝试改为追求更低的目标成本 $a' < a$ 。但是，这

会将我们的网格改为 $\mathscr{V}_{a'}$，而它现在具有不同的基数，因此同样的逻辑也适用于 a'。如果网格恰好过大，类似的推理也同样适用。因此，直观上，我们希望选择 a，使得 $\mathscr{V}_a$ 具有恰当的基数，从而使成本 a 落到我们头上。由此，我们在此意义上寻求某种不动点。

3.2 下界

下界的构造比最初预想的更为微妙。为说明其中的微妙之处，假设我们有一个候选自适应速率 $s \mapsto \epsilon(s)$。我们很容易想到以下下界表述：对于任意 $\eta \in (0,1)$，存在 $c_\eta > 0$，使得对于任意 $0 < c < c_\eta$,

$$\inf_{\varphi} \left\{ P_0\{\varphi \neq 0\} + \max_{1 \le s \le p} \sup_{\substack{f \in \mathcal{F}_s, \\ \|f\|_2 \ge c\epsilon(s)}} P_f\{\varphi \neq 1\} \right\} \ge 1 - \eta.$$

尽管这一判据极其自然，但它存在一个问题。特别是，如果存在 $\bar{s}$，使得 $\epsilon(\bar{s}) \asymp \varepsilon^*(\bar{s})$，其中 $s \mapsto \varepsilon^*(s)$ 表示极小极大速率，那么上述判据便会自动成立。只需对所有 s 取最大值，并选择 $s = \bar{s}$，然后应用极小极大下界即可。说得夸张一点，候选速率 $\epsilon(1) = \varepsilon^*(1)$ 和 $\epsilon(s) = \infty$ 对于 $s \ge 2$ 满足上述判据。此外，从上界的角度来看，存在一个能达到该候选速率的平凡检验。这个下界判据所导致的荒谬结论，将迫使我们认定该候选速率就是自适应检验率。

为避免此类荒谬情况，我们采用另一种下界判据。表述中的关键问题在于：在第二类错误的最大化定义域中，不应包含任何使候选速率与极小极大速率同阶的 s。

定义6。 *Suppose $s \mapsto \varepsilon_{adapt}(p,s,n)$ is a candidate adaptive rate and $\{\mathcal{S}_{p,n}\}_{p,\ n\in\mathbb{N}}$ are a collection of sets satisfying $\mathcal{S}_{p,\ n} \subset [p]$ for all p and n. We say $s \mapsto \varepsilon_{adapt}(p,s,n)$ satisfies the adaptive lower bound criterion with respect to $\{\mathcal{S}_{p,n}\}_{p,\ n\in\mathbb{N}}$ if the following two conditions hold.*

(i) For any $\eta \in (0,1)$ there exists $c_\eta > 0$ depending only on η such that for all $0 < c < c_\eta$,

$$\inf_{\varphi} \left\{ P_0\{\varphi \neq 0\} + \max_{s \in \mathcal{S}_{p,n}} \sup_{\substack{f \in \mathcal{F}_s, \\ \|f\|_2 \ge c\varepsilon_{adapt}(s)}} P_f\{\varphi \neq 1\} \right\} \ge 1 - \eta$$

for all p and n.

(ii) Either

$$\max_{p,n\in\mathbb{N}} \min_{s \in \mathcal{S}_{p,n}} \frac{\varepsilon_{adapt}(p,s,n)}{\varepsilon^*(p,s,n)} = \infty \tag{17}$$

or

$$\max_{p,n\in\mathbb{N}} \max_{s \in [p]} \frac{\varepsilon_{adapt}(p,s,n)}{\varepsilon^*(p,s,n)} < \infty \tag{18}$$

where $s \mapsto \varepsilon^*(p,s,n)$ denotes the minimax separation rate.

直观地说，只有当所选参考集合中不存在任何稀疏度水平，使得候选速率与极小极大速率具有相同阶时，这一判据才是非平凡的。更明确地说，不存在序列 $\{s_{p,n}\}_{p,\ n\in\mathbb{N}}$ 满足 $s_{p,\ n} \in \mathcal{S}_{p,\ n}$ 使得候选速率与极小极大速率沿该序列相匹配（相差至多一个常数因子）。这一判据避免了诸如前文所述的平凡情形。在定义 6 中，对“阶”这一概念进行形式化需要用序列来表述，尽管乍看之下这可能显得不熟悉且不够简洁。定义 6 并非新概念，而是将估计情形中
adaptive rate of convergence on the scale of classes 的概念直接移植到检验语境下，该概念见于 [44]（另请参见 [11]，其中将这一概念应用于稀疏高斯序列模型中的线性泛函估计）。

条件（18）只是要求候选速率与极小极大速率一致，前提是（17）不成立

成立。这样一来，该定义允许极小极大速率成为自适应速率（在无需支付自适应代价的情况下）。

现在我们针对这一判据给出一个下界。定义候选速率

$$\psi_{\mathrm{adapt}}(p,s,n)^2 := \begin{cases} \frac{s}{n}\log\left(1+\frac{p\mathscr{A}_{\mathcal{H}}}{s^2}\right) \vee s\Gamma_{\mathcal{H}}(s,\mathscr{A}_{\mathcal{H}}) & \text{if } s < \sqrt{p\mathscr{A}_{\mathcal{H}}}, \\ s\Gamma_{\mathcal{H}}(s,\mathscr{A}_{\mathcal{H}}) & \text{if } s \geq \sqrt{p\mathscr{A}_{\mathcal{H}}}. \end{cases} \tag{19}$$

这里， $\mathscr{A}_{\mathcal{H}}$ 定义于（15）。考察这一候选速率时，可能会遇到关于$s < (p\mathscr{A}_{\mathcal{H}})^{1/2-\delta}$ 的荒谬情况，因为 $\log\left(1+\frac{p\mathscr{A}_{\mathcal{H}}}{s^2}\right) \asymp \log p$，这意味着候选速率可能与极小极大速率相匹配。这里我们利用了 $\mathscr{A}_{\mathcal{H}}$ 至多随 p 对数增长这一事实。因此，我们取$\mathcal{S} = \left\{s \in [p] : s \geq \sqrt{p\mathscr{A}_{\mathcal{H}}}\right\}$ （为便于记号，省略下标并写作$\mathcal{S} = \mathcal{S}_{p,\ n}$ ）。

由 $\Gamma_{\mathcal{H}}$（s，a）关于 a 递增这一事实可知，（17）或（18）通常应成立。为说明 $\Gamma_{\mathcal{H}}$（s，a）确实关于 a递增，我们固定 $a \leq a'$。注意，对每个 $\nu \in \mathbb{N}$，都有

$$\mu_\nu \wedge \frac{\sqrt{\nu \log\left(1+\frac{pa}{s^2}\right)}}{n} \leq \mu_\nu \wedge \frac{\sqrt{\nu \log\left(1+\frac{pa'}{s^2}\right)}}{n}.$$

对两边关于 ν取最大值得到 $\Gamma_{\mathcal{H}}$（s，a）$\leq \Gamma_{\mathcal{H}}$（$s$，$a'$），即 $\Gamma_{\mathcal{H}}$关于 a递增。现在，考虑 $a = 1$ 和 $a' = \mathscr{A}_{\mathcal{H}}$，以比较候选速率 ψ_{adapt} 与极小极大速率 ε^*。由于 $s \geq \sqrt{p\mathscr{A}_{\mathcal{H}}}$，我们有 $\log\left(1+\frac{p}{s^2}\right) \asymp \frac{p}{s^2}$和 $\log\left(1+\frac{p\mathscr{A}_{\mathcal{H}}}{s^2}\right) \asymp \frac{p\mathscr{A}_{\mathcal{H}}}{s^2}$。根据上述表达式，直观上，在典型的非病态情形下，(17) 或 (18) 应成立；可以针对具体问题轻松检验这些条件（例如第 3.4节中的问题）。

为陈述下界，需要以下集合

$$\widetilde{\mathscr{V}}_{\mathcal{H}} := \left\{2^k : k \in \mathbb{N}\cup\{0\} \text{ and } 2^{k-1} < \nu_{\mathcal{H}}(s,\mathscr{A}_{\mathcal{H}}) \leq 2^k \text{ for some } s \in \mathcal{S}\right\}. \tag{20}$$

定义了 $\widetilde{\mathscr{V}}_{\mathcal{H}}$ 之后，我们就可以陈述下界了。下面的结果针对 6 中定义的 ψ_{adapt} 由（19）给出，且相对于我们选择的 $\mathcal{S} = \left\{s \in [p]$ 建立了定义 $s \geq \sqrt{p\mathscr{A}_{\mathcal{H}}}\right\}$（i）的条件。

定理 2. *Suppose* $\mathscr{A}_{\mathcal{H}} \asymp \log\left(e|\widetilde{\mathscr{V}}_{\mathcal{H}}|\right)$. *Assume further* $\log(1+p\mathscr{A}_{\mathcal{H}}) \leq \frac{n}{2}$. *If* $\eta \in (0,1)$ *then there exists* $c_\eta > 0$ *depending only on* η *such that for all* $0 < c < c_\eta$, *we have*

$$\inf_{\varphi}\left\{P_0\{\varphi \neq 0\} + \max_{s\in\mathcal{S}} \sup_{\substack{f\in\mathcal{F}_s,\\ \|f\|_2 \geq c\psi_{adapt}}} P_f\{\varphi \neq 1\}\right\} \geq 1-\eta$$

where ψ_{adapt} *is given by (19) and* $\mathcal{S} = \left\{s \in [p] : s \geq \sqrt{p\mathscr{A}_{\mathcal{H}}}\right\}$.

该条件 $\mathscr{A}_{\mathcal{H}} \asymp \log\left(e|\widetilde{\mathscr{V}}_{\mathcal{H}}|\right)$ 较为宽松；例如，当特征值满足如下多项式衰减时，该条件成立：$\mu_k = c_\alpha k^{-2\alpha}$ 其中 $\alpha > 0$，且正的通用常数 c_α。当存在如下指数衰减时，该条件也成立：$\mu_k = c_1 e^{-c_2 j^\gamma}$ 其中 $\gamma > 0$，且正的通用常数 c_1 和 c_2。第 3.4 节进一步讨论了这两种情形。

证明策略从总体上看，与第 2.3 节中的策略相同，但增加了一项复杂性：先验还应当对稀疏度进行随机化，以刻画未知稀疏度所带来的额外困难。随机稀疏度通过随机抽取一个截断点 ν 并提取由此诱导的稀疏度来生成。先验的其余构造与第 2.3节中的构造类似，但随机稀疏度使得分析更加复杂。

最后请注意，定理 2 说明了在稀疏度水平$\mathcal{S} = \left\{s \in [p] : s \geq \sqrt{p\mathscr{A}_{\mathcal{H}}}\right\}$下，为实现自适应所付出的代价。对于更小的稀疏度，定理并未作出任何论断。值得研究候选项 ψ_{adapt} 是否也刻画了满足以下条件的稀疏度水平 s $(p\mathscr{A}_{\mathcal{H}})^{1/2-\delta} < s < \sqrt{p\mathscr{A}_{\mathcal{H}}}$ 对所有 $\delta \in (0,1/2]$（例如 $s \asymp \sqrt{p\mathscr{A}_{\mathcal{H}}}/\log(p\mathscr{A}_{\mathcal{H}})$）下的代价；这个问题留待未来研究。

3.3 上界

本节将构造一个自适应检验。回顾式 $\mathscr{A}_{\mathcal{H}}$ 和 $\mathscr{V}_{\mathcal{H}}$，它们分别见于（15）和（16）。定义

$$\tau^2_{\text{adapt}}(p,s,n)=\begin{cases}\left(\frac{s}{n}\log\left(1+\frac{p\mathscr{A}_{\mathcal{H}}}{s^2}\right)\right)\vee\left(\frac{s}{n}\sqrt{\nu_{\mathcal{H}}(s,\mathscr{A}_{\mathcal{H}})\log\left(1+\frac{p\mathscr{A}_{\mathcal{H}}}{s^2}\right)}\right) & \text{if } s<\sqrt{p\mathscr{A}_{\mathcal{H}}},\\ \frac{s}{n}\sqrt{\nu_{\mathcal{H}}(s,\mathscr{A}_{\mathcal{H}})\log\left(1+\frac{p\mathscr{A}_{\mathcal{H}}}{s^2}\right)} & \text{if } s\geq\sqrt{p\mathscr{A}_{\mathcal{H}}}.\end{cases}\tag{21}$$

在该检验中，使用以下几何网格集合。定义

$$\mathscr{S}:=\left\{1,2,4,...,2^{\left\lceil\log_2\sqrt{p\mathscr{A}_{\mathcal{H}}}\right\rceil-1}\right\}\cup\{p\}.\tag{22}$$

选择将 $\mathscr{S}$ 作为几何网格在统计上并非关键，但这样做在计算上更有优势。

定理 3. *There exist universal positive constants D, K_2, K_2', and K_3 such that the following holds. Let $\mathscr{A}_{\mathcal{H}}$, $\mathscr{V}_{\mathcal{H}}$, and $\mathscr{S}$ be given by (15), (16), and (22) respectively. For $\nu\in\mathscr{V}_{\mathcal{H}}$ and $s\in[p]$, set $d_\nu=\nu\vee\lceil D\rceil$ and set*

$$r_{\nu,s}:=K_2\left(d_\nu\log\left(1+\frac{p\mathscr{A}_{\mathcal{H}}}{s^2}\right)\right)^{1/4},$$
$$r'_s:=K_2'\sqrt{\log\left(1+\frac{p\mathscr{A}_{\mathcal{H}}}{s^2}\right)}.$$

If $\eta\in(0,1)$ then there exists $C_\eta>0$ depending only on η such that for all $C>C_\eta$, we have

$$P_0\left\{\max_{\nu\in\mathscr{V}_{\mathcal{H}}}\max_{s\in\mathscr{S}}\varphi_{\nu,s}\neq 0\right\}+\max_{1\leq s^*\leq p}\sup_{\substack{f\in\mathcal{F}_{s^*},\\ \|f\|_2\geq C\tau_{adapt}(p,s^*,n)}}P_f\left\{\max_{\nu\in\mathscr{V}_{\mathcal{H}}}\max_{s\in\mathscr{S}}\varphi_{\nu,s}\neq 1\right\}\leq\eta$$

where τ_{adapt} is given by (21). Here, $\varphi_{\nu,\ s}$ is given by

$$\varphi_{\nu,s}:=\begin{cases}\mathbb{I}\left\{T_{r_{\nu,s}}(d_\nu)>Cs\sqrt{\nu\log\left(1+\frac{p\mathscr{A}_{\mathcal{H}}}{s^2}\right)}\right\} & \text{if } s<\sqrt{p\mathscr{A}_{\mathcal{H}}} \text{ and } \sqrt{\log\left(1+\frac{p\mathscr{A}_{\mathcal{H}}}{s^2}\right)}\leq K_3\sqrt{d_\nu},\\ \mathbb{I}\left\{T_{r'_s}(d_\nu)>Cs\log\left(1+\frac{p\mathscr{A}_{\mathcal{H}}}{s^2}\right)\right\} & \text{if } s<\sqrt{p\mathscr{A}_{\mathcal{H}}} \text{ and } \sqrt{\log\left(1+\frac{p\mathscr{A}_{\mathcal{H}}}{s^2}\right)}>K_3\sqrt{d_\nu},\\ \mathbb{I}\left\{n\sum_{j=1}^p\sum_{k\leq\nu}X^2_{k,j}>\nu p+C\sqrt{\nu p\mathscr{A}_{\mathcal{H}}}\right\} & \text{if } s\geq\sqrt{p\mathscr{A}_{\mathcal{H}}}\end{cases}$$

and the statistic $T_r(d)$ is given by (12).

如前所述，自适应检验需要扫描几何网格 $\mathscr{V}_{\mathcal{H}}$ 中 ν 的可能取值。因此，需要付出与 $\mathscr{A}_{\mathcal{H}}$ 相关的代价。注意，对于 $s<\sqrt{p\mathscr{A}_{\mathcal{H}}}$，我们还需要扫描 $s\in\mathscr{S}$，以便为统计量 $T_r(d)$ 正确设置阈值。事实证明，这种额外扫描不会带来代价；代价来自需要扫描 ν。

要理解扫描 ν 可能带来的代价，最直观的方式是考虑在扫描 χ^2 统计量（其自由度各不相同）时控制第一类错误的必要性。粗略地说，有必要选择阈值 t_ν，使得第一类错误 $P\left(\bigcup_{\nu\in\mathscr{V}_{\mathcal{H}}}\{\chi^2_{\nu p}-\nu p>t_\nu\}\right)$ 小于某个预先规定的错误水平。根据并集界和（14），考虑尾界 $P\left(\bigcup_{\nu\in\mathscr{V}_{\mathcal{H}}}\{\chi^2_{\nu p}-\nu p>2\sqrt{\nu pu}+2u\}\right)\leq\sum_{\nu\in\mathscr{V}_{\mathcal{H}}}P\{\chi^2_{\nu p}-\nu p>2\sqrt{\nu pu}+2u\}\trianglelefteq|\mathscr{V}_{\mathcal{H}}|e^{-u}$。为了确保第一类错误较小，该尾界表明应选择 $t_\nu\asymp\sqrt{\nu p}\log|\mathscr{V}_{\mathcal{H}}|+2\log|\mathscr{V}_{\mathcal{H}}|\asymp\sqrt{\nu p\log|\mathscr{V}_{\mathcal{H}}|}$，其中后一个等价关系源于以下事实：$|\mathscr{V}_{\mathcal{H}}|$ 的增长至多是 p 的对数级别。由于阈值 t_ν 被膨胀了，信号也必须具有更大的范数才能被检测到。对数膨胀因子是 χ^2 尾部。

3.4 特殊情形

为了说明适应代价如何微妙地取决于函数空间 $\mathcal{H}$ 的选择，我们考虑多种不同的情形。在全文中，均假设 $\log(1+p\mathscr{A}_{\mathcal{H}}) \leq n/2$ as，同时满足我们结果所需的其他假设。

3.4.1 索об列夫

以 $\mu_k \asymp k^{-2\alpha}$ 作为光滑性为 α 的 Sobolev空间代表，可以证明 $\mathscr{A}_{\mathcal{H}} \asymp \log(e|\widetilde{\mathscr{V}}_{\mathcal{H}}|) \asymp \log\log p$ (回顾（15）和（20））。我们的上界和下界表明，自适应检验率为

$$\varepsilon_{\text{adapt}}(p,s,n)^2 \asymp \begin{cases} \frac{s\log\left(\frac{p\log\log p}{s^2}\right)}{n} + s\left(\frac{n}{\sqrt{\log\left(\frac{p\log\log p}{s^2}\right)}}\right)^{-\frac{4\alpha}{4\alpha+1}} & \text{if } s < \sqrt{p\log\log p}, \\ s\left(\frac{ns}{\sqrt{p\log\log p}}\right)^{-\frac{4\alpha}{4\alpha+1}} & \text{if } s \geq \sqrt{p\log\log p}. \end{cases}$$

Ingster 和 Lepski [25] 分别考虑稀疏区域 $s = O(p^{1/2-\delta})$ 和稠密区域$s = p^{1/2+\delta}$，其中 $\delta \in (0, 1/2)$。在稀疏区域，他们构造了一种能够达到极小极大速率的自适应检验，无需付出任何代价。在稠密区域，他们不仅给出了一个达到 $s\left(\frac{ns}{\sqrt{p\log\log p}}\right)^{-\frac{4\alpha}{4\alpha+1}}$ 的检验，还给出了一个下界（现在用我们的下界记号表示，可取 $\mathcal{S} = [p]$），表明该速率无法进一步改进。如上式所示，我们的检验在这些情形下具有最优性。

读者在将[25] 与我们的结果进行比较时应谨慎。在我们的设定中，未知稀疏度可以在整个范围 $s \in [p]$内变化。相比之下， [25] 考虑了两种不同情形：未知稀疏度 s 要么被限制在 $s \leq p^{1/2-\delta}$，要么被限制在 $s \geq p^{1/2+\delta}$。具体而言， s 的确切值未知，但 [25] 假设统计学家知道当前属于这两种情形中的哪一种。相比之下，我们不作这样的假设；稀疏度完全未知。我们认为，这种情形更贴近实际工作者面临的问题。尽管我们的检验针对看似更困难的设定构造，结果表明，在 Ingster 和 Lepski 的设定下，它仍然是最优的。

我们的结果对 $\sqrt{p}$ 提供了更细致的刻画，而 [25]未能做到。在他们文章的定理 3 中，Ingster 和 Lepski 提出了一种自适应检验程序，适用于情形$s \asymp \sqrt{p}$。该程序要求信号强度的平方阶至少达到

$$\sqrt{p}\left(\frac{n}{\sqrt{\log\log p}}\right)^{-\frac{4\alpha}{4\alpha+1}}.$$

这是次优的，因为我们的检验达到了

$$\begin{cases} \sqrt{p}\left(\frac{n}{\sqrt{\log\log\log p}}\right)^{-\frac{4\alpha}{4\alpha+1}} & \text{if } \log\log\log p \leq n^{\frac{1}{2\alpha+1}}, \\ \frac{\sqrt{p}\log\log\log p}{n} & \text{if } \log\log\log p > n^{\frac{1}{2\alpha+1}}, \end{cases}$$

其速度更快。

3.4.2 有限维

考虑有限维结构 $1 = \mu_1 = \mu_2 = \ldots = \mu_m > \mu_{m+1} = \ldots = 0$ 其中 m 是一个正整数。如果 $m < n^2/\log(1+p)$，可以证明 $\mathscr{A}_{\mathcal{H}} \asymp 1$，因此我们的自适应检验可以达到极小极大分离率。换言之，自适应无需付出代价。

3.4.3 指数衰减

考虑特征值 $\mu_k = c_1 e^{-c_2 k^\gamma}$ 呈指数衰减的情形，其中 $\gamma > 0$。可以证明 $\sim \mathscr{A}_{\mathcal{H}} \asymp \log(e|\mathscr{V}_{\mathcal{H}}|) \asymp \log\log\log p$，因此自适应速率为

$$\varepsilon_{\mathrm{adapt}}(p,s,n)^2 \asymp \begin{cases} \frac{s\log\left(\frac{p\log\log\log p}{s^2}\right)}{n} + s\log^{\frac{1}{2\gamma}}\left(\frac{n}{\sqrt{\log\left(\frac{p\log\log\log p}{s^2}\right)}}\right)\cdot\frac{\sqrt{\log\left(\frac{p\log\log\log p}{s^2}\right)}}{n} & \text{if } s < \sqrt{p\log\log\log p}, \\ \frac{\sqrt{p\log\log\log p}}{n}\log^{\frac{1}{2\gamma}}\left(\frac{ns}{\sqrt{p\log\log\log p}}\right) & \text{if } s \geq \sqrt{p\log\log\log p}. \end{cases}$$

此处的自适应代价增长非常缓慢，这或许也表明基于高斯核的再生核希尔伯特空间相对“较小”（如第 2.5 节所述）。

4 对稀疏性和光滑性的双重自适应

到目前为止，我们一直假设空间 $\mathcal{H}$ 已知。然而，$\mathcal{H}$ 很可能是由某个超参数（如光滑度水平）索引的一组空间中的一个。通常，除了稀疏性未知之外，该超参数的真实取值也未知，因此我们希望了解检验速率如何变化。为避免过于抽象，我们遵循 [17, 25]，采用一个工作模型，其中 $\mathscr{T}_s$ 具有 μ_k 如下形式：$\mu_k = k^{-2\alpha}$ 这是具有光滑性 α的索об列夫空间的典型形式。为强调对 α的依赖，我们记 $\mathscr{T}(s,\ \alpha)$ 及其对应的空间 $\mathcal{F}(s,\ \alpha)$。重申一下，我们关注如何同时对稀疏度 s和光滑度水平 α进行自适应。

Ingster 和 Lepski [25] 研究同时针对稀疏性和光滑性的自适应问题。特别是，他们研究在端点已知的闭区间内，未知 $\alpha \in [\alpha_0, \alpha_1]$ 的自适应问题。正如 $\alpha_0 < \alpha_1$ 中所论证的，[25], 由于 α_1 已知，第 1.2.2 节中任意$\alpha \in [\alpha_0, \alpha_1]$ 的序列空间基都可以取为 α_1-正则，因此统计学家知道该基。此外，他们分别研究稠密区域 $s \geq p^{1/2+\delta}$ 和稀疏区域 $s < p^{1/2-\delta}$，其中 $\delta \in (0,\ 1/2)$。我们采用与他们相同的设定。

Ingster 和 Lepski [25] 声称以下自适应速率。在稠密情形 $s \geq p^{1/2+\delta}$ 下，他们作出假设 $\frac{p}{s^2}\log\left(\frac{n^2s^2/p}{n^2s^2/p\log()}\right) \to 0$，并给出如下结论（[25] 中的定理 4）：

$$s\left(\frac{ns}{\sqrt{p\log\left(\frac{ns}{\sqrt{p}}\right)}}\right)^{-\frac{4\alpha}{4\alpha+1}}.$$

相比之下[1], 我们将证明稠密情形下的自适应速率 $s \geq p^{1/2+\delta}$ 为

$$\varepsilon^2_{\mathrm{dense}}(p,s,n,\alpha) \asymp s\left(\frac{ns}{\sqrt{p\log\log(np)}}\right)^{-\frac{4\alpha}{4\alpha+1}}.$$

细心的读者会注意到，在特殊情形下 $p = 1$，我们的结果恢复了 Spokoiny 证明的自适应分离速率 $\left(n/\sqrt{\log\log n}\right)^{-\frac{4\alpha}{4\alpha+1}}$ [41]。在稀疏情形下 $s < p^{1/2-\delta}$，他们声称的速率（[25]中的定理 5）为

$$\frac{s\log p}{n} + s\left(\frac{n}{\sqrt{\log(p\log n)}}\right)^{-\frac{4\alpha}{4\alpha+1}}. \tag{23}$$

作为部分修正，我们证明以下下界

$$\frac{s\log(p\log\log n)}{n} + s\left(\frac{n}{\sqrt{\log(p\log\log n)}}\right)^{-\frac{4\alpha}{4\alpha+1}}. \tag{24}$$

[1]同样的错误也影响了[25]中定理 4 和定理 5 的下界证明。他们定义的先验不以参数空间为支撑。在该先验下（见 [25]中的式（75）），各个坐标函数 f_j 最终具有不同的光滑度级别 α_j，而不是共享同一个 α。

在情形 $\log\ p \gtrsim \log\log\ n$中，(23) 和 (24) 彼此一致，并且与极小极大速率一致（见第 2.5.1节）。然而，在 $\log p \lesssim \log\log\ n$的情况下，(23) 与 (24) 之间可能存在差异。

4.1 稠密

固定 $\delta \in (0, 1/2)$ 和 $\alpha_0 < \alpha_1$。回顾在稠密情形下，$s \geq p^{1/2+\delta}$。定义

$$\tau^2_{\text{dense}}(p, s, n, \alpha) := s\left(\frac{ns}{\sqrt{p\log\log(np)}}\right)^{-\frac{4\alpha}{4\alpha+1}}. \tag{25}$$

如前所述，我们将在稠密情形下证明这一速率。为用于检验程序，定义几何网格

$$\mathcal{V}_{\text{test}} := \left\{1, 2, 4, 8, ...2^{\left\lceil \log_2\left(\left(\frac{np}{\sqrt{p\log\log(np)}}\right)^{\frac{2}{4\alpha_0+1}}\right)\right\rceil}\right\}. \tag{26}$$

请注意，统计学家无需知道 s 或 α 即可构造 $\mathcal{V}_{\text{test}}$。另请注意，$\log|\mathcal{V}_{\text{test}}| \asymp \log\log(np)$。

4.1.1 上界

所采用的检验程序遵循自适应检验中的典型策略，即针对网格 $\mathcal{V}_{\text{test}}$ 中 ν 的可能取值分别构造检验，然后对各个检验取最大值以进行信号检测。自适应的代价涉及 $\mathcal{V}_{\text{test}}$ 的基数的对数（即 $\log\log(np)$）。由于适用的是稠密情形，各个检验均为 χ^2 检验。

定理 4。 *Fix $\delta \in (0, 1/2)$ and $\alpha_0 < \alpha_1$. If $\eta \in (0, 1)$ then there exists $C_\eta > 0$ depending only on η such that for all $C > C_\eta$, we have*

$$P_0\left\{\max_{\nu\in\mathcal{V}_{test}} \varphi_\nu = 1\right\} + \max_{s\geq p^{1/2+\delta}} \sup_{\alpha\in[\alpha_0,\alpha_1]} \sup_{\substack{f\in\mathcal{F}(s,\alpha),\\ ||f||_2\geq C\tau_{dense}(p,s,n,\alpha)}} P_f\left\{\max_{\nu\in\mathcal{V}_{test}} \varphi_\nu = 0\right\} \leq \eta$$

where

$$\varphi_\nu = \mathbb{I}\left\{n\sum_{j=1}^p \sum_{k\leq\nu} X^2_{k,j} \geq \nu p + K_\eta\left(\sqrt{\nu p\log\log(np)} + \log\log(np)\right)\right\}.$$

with K_η being a constant depending only on η. Here τ_{dense} is given by (25) and $\mathcal{V}_{test}$ is given by (26).

4.1.2 下界

下面的定理给出了下界。注意，经过直接修改以纳入对 6 自适应，可知它满足定义 α。具体而言，第 3.2 节所述的荒谬情况可能造成的困难并不会出现，因为候选速率

$$\tau^2_{\text{dense}}(p, s, n, \alpha) = s\left(\frac{ns}{\sqrt{p\log\log(np)}}\right)^{-\frac{4\alpha}{4\alpha+1}}$$

其阶从不与极小极大速率 $\varepsilon^*(p, s, n, \alpha)^2 = s\left(\frac{ns}{\sqrt{p}}\right)^{-\frac{4\alpha}{4\alpha+1}}$ 相同。

定理 5。 *Fix $\delta \in (0, 1/2)$ and $\alpha_0 < \alpha_1$. If $\eta \in (0, 1)$ then there exists $c_\eta > 0$ depending only on η such that for all $0 < c < c_\eta$, we have*

$$\inf_\varphi \left\{P_0\{\varphi \neq 0\} + \max_{s\geq p^{1/2+\delta}} \sup_{\alpha\in[\alpha_0,\alpha_1]} \sup_{\substack{f\in\mathcal{F}(s,\alpha),\\ ||f||_2\geq c\tau_{dense}(p,s,n,\alpha)}} P_f\{\varphi \neq 1\}\right\} \geq 1 - \eta$$

where τ_{dense} is given by (25).

4.2 稀疏

固定 $\delta \in (0, 1/2)$ 和 $\alpha_0 < \alpha_1$。回忆一下，在稀疏区域中，我们有 $s < p^{1/2-\delta}$。定义

$$\tau^2_{\text{sparse}}(p, s, n, \alpha) := \frac{s \log (p \log \log n)}{n} + s \left(\frac{n}{\sqrt{\log(p \log \log n)}} \right)^{-\frac{4\alpha}{4\alpha+1}}. \tag{27}$$

Ingster 和 Lepski [25] 给出了一个达到（23）的检验程序，从而确立了上界。正如前文所述，当 $\log p \gtrsim \operatorname{loglog} n$ 时，不仅（27）和（23）彼此匹配，而且还与极小极大速率相匹配（见第 2.5.1 节）。在 $\log p \lesssim \log \log n$ 的情形下，从下界的角度来看，自适应代价是否不可避免尚不清楚。我们通过给出一个下界来补充他们的上界，该下界表明确实必须付出自适应代价。然而，我们所确定的代价可能并不精确。具体而言，在 p 是一个较大的通用常数但远小于 n 的情况下，稀疏区域速率（27）包含 $\sqrt{\log \log \log n}$。根据 Spokoiny的文章 [41]，$\sqrt{\operatorname{loglog} n}$ 则是预期的结果。我们将精确确定这一代价留待未来研究。

定理 6. *Fix $\delta \in (0, 1/2)$ and $\alpha_0 < \alpha_1$. If $\eta \in (0, 1)$ then there exists $c_\eta > 0$ depending only on η such that for all $0 < c < c_\eta$, we have*

$$\inf_{\varphi} \left\{ P_0 \{\varphi \neq 0\} + \max_{s<p^{1/2-\delta}} \sup_{\alpha \in [\alpha_0, \alpha_1]} \sup_{\substack{f \in \mathcal{F}(s,\alpha), \\ ||f||_2 \geq c\tau_{sparse}(p,s,n,\alpha)}} P_f \{\varphi \neq 1\} \right\} \geq 1 - \eta$$

where τ_{sparse} is given by (27).

如定理 5所述，定理 6 中的下界 6 满足定义，但需做直接修改以纳入对 α的自适应。

5 讨论

5.1 放宽中心化假设

在第 1.2.2 节中，参数空间（4）的定义将信号 f 限制为中心化。文中指出，这一限制较宽松，可以放宽；本节将进一步阐述这一点。

设 f 是一个加性函数（可能未中心化），且每个 $f_j \in \mathcal{H}$。令 $\bar{f} = \int_{[0,1]^p} f(x)\, dx$，并注意到我们可以写成 $f = \bar{f}\mathbf{1}_p + (f - \bar{f}\mathbf{1}_p)$，其中 $\mathbf{1}_p \in L^2([0,1]^p)$ 是恒等于 1 的常函数。由于 f 是一个加性函数，对任意 $x \in [0,1]^p$，有 $f(x) = \bar{f} + g(x)$，其中 $g(x) = \sum_{j=1}^p g_j(x_j)$，且 $g_j(x_j) = f_j(x_j) - \int_0^1 f_j(t)\, dt$。换言之，我们有 $f = \bar{f}\mathbf{1}_p + g$，其中 g 本身是一个加性函数，且每个 $g_j \in \mathcal{H}_0$。

断言 f 是一个稀疏加性函数，意味着 g 是一个（中心化的）稀疏加性函数。按照 [17, 38], ，我们对中心化分量函数施加约束（即再生核希尔伯特空间范数约束） g_j。因此，可以考虑以下非中心化信号检测问题：

$$H_0 : f \equiv 0, \tag{28}$$

$$H_1 : ||f||_2 \geq \varepsilon,\ f \text{ is } s\text{-sparse additive, and } g \in \mathcal{F}_s \tag{29}$$

其中 $\mathcal{F}_s$ 由（4）给出。令 ε^* 表示极小极大分离率。由正交性，考虑 $||f||_2^2 = \bar{f}^2 + ||g||_2^2$。因此，存在两个相关的检测问题，即问题 I

$$H_0 : f \equiv 0,$$

$$H_1 : |\bar{f}| \geq \varepsilon_1 \text{ and } f \text{ is } s\text{-sparse additive}$$

和问题 II

$$H_0 : f \equiv 0,$$
$$H_1 : ||g||_2 \geq \varepsilon_2 \text{ and } g \in \mathcal{F}_s.$$

令 ε_1^* 和 ε_2^* 分别表示相应问题的极小极大分离率。我们声称 $(\varepsilon^*)^2 \asymp (\varepsilon_1^*)^2 + (\varepsilon_2^*)^2$。由 $||f||^2 = \bar{f}^2 + ||g||_2^2 \geq \bar{f}^2 \vee ||g||_2^2$ 可直接得到 $(\varepsilon^*)^2 \gtrsim (\varepsilon_1^*)^2 \vee (\varepsilon_2^*)^2 \asymp (\varepsilon_1^*)^2 + (\varepsilon_2^*)^2$, , 因此下界得证。为了证明上界，考虑若 $||f||_2^2 \geq C\left((\varepsilon_1^*)^2 + (\varepsilon_2^*)^2\right)$，则根据三角不等式可知，以下两者之一成立： $|\bar{f}|^2 \geq C(\varepsilon_1^*)^2$ 或 $||g||_2^2 \geq C(\varepsilon_2^*)^2$。因此，信号足以被检测到，并且在各子问题的极小极大检验中取最大值的检验是最优的。因此，上界 $(\varepsilon^*)^2 \lesssim (\varepsilon_1^*)^2 + (\varepsilon_2^*)^2$ 也得证。

剩下的就是确定极小极大速率 ε_1^* 和 ε_2^*。我们从零开始编号，并令 $\{\psi_k\}_{k=0}^{\infty}$ 和 $\{\mu_k\}_{k=0}^{\infty}$ 分别为再生核希尔伯特空间 $\mathcal{H}_0$对应的特征函数和特征值。回忆一下，$\{\psi_k\}_{k=0}^{\infty}$ 在通常的 L^2 内积下构成 $L^2([0,1])$ 的标准正交基。根据定义， $\mathcal{H}_0$ 与 $L^2([0,1])$ 中常函数所张成的空间正交，因此不失一般性，我们可以令 ψ_0 为取值恒为一的常函数， $\mu_0 = 0$，并令 $\{\psi_k\}_{k=1}^{\infty}$ 为其余与常函数正交（在 L^2 下）的特征函数。

数据 $\{X_{k,j}\}_{k\in\mathbb{N},\ j\in[p]}$ (定义见 (5)) 对问题 II 是充分的。因此，极小极大速率ε_2^* 恰由本文建立的结果给出。极小极大速率 ε_1^* 可以按以下方式给出上界。考虑到 $\langle \mathbf{1}_p, dY_x \rangle_{L^2([0,1]^p)} \sim N(\bar{f}, \frac{1}{n})$，因为 $||\mathbf{1}_p||_2^2 = 1$。因此，可以达到参数速率 $\varepsilon_1^* \lesssim \frac{1}{\sqrt{n}}$。于是，$(\varepsilon^*)^2 \asymp (\varepsilon_1^*)^2 + (\varepsilon_2^*)^2 \asymp (\varepsilon_2^*)^2$。换言之，中心化与信号检测问题基本极限无关。

5.2 尖锐常数

正如我们先前指出的，我们的结果并未涉及检测边界中的尖锐常数。如何精确刻画检测边界中的常数，是一个有趣且可能颇为微妙的问题。在渐近设置下，对于索ob列夫情形，Gayraud 和 Ingster [17] 成功推导出了稀疏区域中的尖锐常数 $s = p^{1-\beta}$ 对于固定的$\beta \in (1/2,1)$，在满足条件 $\log\ p = o(n^{\frac{1}{2\alpha+1}})$ 的情况下。Gayraud 和 Ingster 指出，这个条件实际上是必不可少的：违反该条件时，检测边界不再依赖于 β。用我们的记号，这个条件可以简洁地表述为与以下条件等价:

$$\frac{\log(p/s^2)}{n} = o\left(\Gamma_{\mathcal{H}}\right).$$

在索ob列夫情形下，这种对应关系表明，对于一般的$\mathcal{H}$，这个条件可能确实是必不可少的。若果真如此，弄清这一点并在条件成立时推导出尖锐常数，将会很有意思。

需要明确的是，我们提出的程序在常数方面可能是次优的（甚至很可能如此）。事实上，关于稀疏信号检测的现有文献，无论是在稀疏序列模型 [14, 18] 还是稀疏线性回归 [1, 26] 中，都依赖高阶检验型检验来达到最优常数。Gayraud 和 Ingster [17] 本人也使用高阶批评。对于一般空间$\mathcal{H}$，不应将我们的程序视为唯一速率最优的检验。在稀疏性情形 $s = p^{1/2-\delta}$下，我们推测，若采用一种考虑核函数特征结构的类似高阶批评型统计量，或许不仅能达到最优速率，还能达到尖锐常数。

5.3 未来方向

未来的研究有多个方向。首先，我们只考虑了一个空间 $\mathcal{H}$，所有分量函数 f_j 都属于该空间。在某些情形下，针对每个分量分别考虑不同的空间$\mathcal{H}_j$ 可能更为理想。Raskutti 等人 [38] 在考虑多个核时（在对核函数作出某些条件限制的情况下）获得了极小极大估计率。我们设想，本文中的广义方法可以扩展到允许使用多个核的情形，以确定极小极大分离率。与其使用一个

共同的 $\nu_{\mathcal{H}}$，不同坐标可能需要使用不同的量$\nu_{\mathcal{H}_j}$，并且可以以自然的方式修改检验统计量。这里建立的理论可以直接使用，而且极小极大分离率似乎可以通过直接的方式得到。

另一条研究路径是考虑以 ℓ_q 约束形式表示的$0 < q < 1$“软”稀疏性。Yuan和Zhou[50] 针对特征值呈多项式衰减的再生核希尔伯特空间（例如 Sobolev空间）推导了极小极大估计率。在信号检测方面，可以设想，将高斯序列模型中具有 ℓ_q 均值界的二次泛函估计量加以扩展，并用作检验统计量 [10]。研究[10] 中的硬稀疏性和软稀疏性设定，处理方式颇为相似。要获得软稀疏性下的极小极大分离率，可能无需发展太多额外理论。

由于在许多问题中，假设检验与泛函估计密切相关，因此自然会想到研究稀疏加性模型背景下的泛函估计。例如，估计二次泛函 $||f||_2^2$ 或范数 $||f||_2$会很有意义。非参数文献中众所周知，对于奇数 r，估计 L_r 范数所得的速率与检验以及偶数 r [19, 32]的速率有显著差异。一个值得深入探索的方向是在稀疏加性模型设定下研究同一问题。

此外，考虑非参数回归模型 $Y_i = f(X_i) + Z_i$，其中设计分布 $X_i \overset{iid}{\sim} P_X$ 在坐标之间表现出某种依赖关系，将会很有意义。我们所依赖的白噪声模型（1）与非参数回归模型之间的对应关系要求设计分布在 $[0,1]^p$ 上接近均匀分布。然而，在实际情况下，X_i 的坐标通常存在依赖关系，因此研究这种依赖关系如何影响检验的基本极限将会很有意义。

最后，与前文讨论的方向有所不同，研究其他模型中组稀疏性下的信号检测问题也很有意义。正如我们所鼓励的那样，我们的结果可以完全从序列空间的角度进行解释，即参数空间为（7）-（8）且参数空间为（6）的问题。从这一角度来看，组稀疏结构显而易见。估计在组稀疏设置下已经得到了广泛研究，尤其是在线性回归中（参见 [47] 及其中的参考文献）。假设检验尚未获得近乎同等程度的研究关注，因此仍有充足的研究空间。我们认为，本文所建立的某些速率特征是组稀疏结构的普遍特征，发现其中的共性将十分引人入胜。

6 致谢

SK 感谢 Oleg Lepski 的亲切通信。SK 的研究部分得到了美国国家科学基金会资助项目 DMS-1547396 和 ECCS-2216912 的支持。CG 的研究部分得到了美国国家科学基金会资助项目 ECCS-2216912、美国国家科学基金会职业奖 DMS-1847590 以及阿尔弗雷德·斯隆研究奖学金的支持。

7 证明

7.1 极小极大上界

7.1.1 稀疏

Proof of Proposition 2. 固定 $\eta \in (0,1)$。我们将在证明结束时设定 C_η，因此目前令 $C > C_\eta$。通用常数 K_1 将在证明过程中选取。令 L^* 表示引理21中的通用常数。设 $K_2 := 1 \vee {}^{1}{}_{(\log 2)}$ $^{1/4} \vee c^{-1/4} \vee \left(\frac{L^*}{c}\right)^{1/4}$ 以及 $K_3 := \frac{L^*}{K_2^2}$其中 ~~$c = c^* \wedge c^{**}$~~ 其中 c^* 和 c^{**} 分别是引理15 和14中指数项里的通用常数。

我们首先对第一类错误进行界定。由于 $\sqrt{\log\left(1+\frac{p}{s^2}\right)} \le K_3\sqrt{d}$，我们有 $1 \le K_2^2\sqrt{\log\left(1+\frac{p}{s^2}\right)} \le$

$K_2^2 K_3\sqrt{d} \le L^*\sqrt{d}$。因此，我们可以应用引理 15 得到

$$P_0\left\{T_r(d) > C^*\left(\sqrt{xpr^4 e^{-\frac{c^* r^4}{d}}} + \frac{d}{r^2}x\right)\right\} \le e^{-x}$$

对任意 $x > 0$。这里， C^* 是一个通用常数。取 $x = C$ ，并注意到只要我们选择 $C_\eta \ge 1$，就有 $C > 1$ ，于是可得

$$\begin{aligned}
C^*\left(\sqrt{xpr^4 e^{-\frac{c^* r^4}{d}}} + \frac{d}{r^2}x\right) &\le C^* C\left(K_2^2\sqrt{pd\log\left(1+\frac{p}{s^2}\right)e^{-c^* K_2^4 \log\left(1+\frac{p}{s^2}\right)}} + \frac{d}{r^2}\right) \\
&\le C^* C\left(K_2^2\sqrt{d\log\left(1+\frac{p}{s^2}\right)}\sqrt{p\cdot\frac{s^2}{s^2+p}} + \frac{d}{r^2}\right) \\
&\le C^* C\left(K_2^2 s\sqrt{d\log\left(1+\frac{p}{s^2}\right)} + \frac{d}{r^2}\right) \\
&\le 2C^* K_2^2 C s\sqrt{d\log\left(1+\frac{p}{s^2}\right)} \\
&= C K_1 n\tau(p,s,n)^2
\end{aligned}$$

这里我们用到了 $c^* K_2^4 \ge 1$ ，并选择了 $K_1 = 2C^* K_2^2$。注意，我们还用到了 $\frac{d}{r^2} \le \sqrt{d} \le K_2^2 s\sqrt{d\log\left(1+\frac{p}{s^2}\right)}$。因此，取 K_1、 K_2和 x，可得

$$P_0\left\{T_r(d) > C K_1 n\tau(p,s,n)^2\right\} \le e^{-x} = e^{-C} \le e^{-C_\eta} \le \frac{\eta}{2}$$

只要我们选择 $C_\eta \ge \log\left(\frac{2}{\eta}\right) \vee 1$。

现在我们考察第二类错误。为了界定第二类错误，我们将使用切比雪夫不等式。具体而言，考虑任意满足 $||f||_2 \ge C\tau(p,s,n)$ 的 $f \in \mathcal{F}_s$，有

$$\begin{aligned}
P_f\left\{T_r(d) \le C K_1 n\tau(p,s,n)^2\right\} &= P_f\left\{E_f(T_r(d)) - C K_1 n\tau(p,s,n)^2 \le E_f(T_r(d)) - T_r(d)\right\} \\
&\le \frac{\mathrm{Var}_f\left(T_r(d)\right)}{\left(E_f(T_r(d)) - C K_1 n\tau(p,s,n)^2\right)^2} \qquad (30)
\end{aligned}$$

只要 $E_f(T_r(d)) \ge C K_1 n\tau(p,s,n)^2$。为确保切比雪夫不等式的应用有效，我们必须计算 $T_r(d)$ 的期望和方差的合适界，以下引理将给出这些界。

引理 4. *If C_η is larger than some sufficiently large universal constant, then $E_f(T_r(d)) \ge 2C K_1 n\tau(p,s,n)^2$.*

引理 5. *If C_η is larger than some sufficiently large universal constant, then 方差$_f(T_r(d)) \le C^\dagger\left(s^2 r^4 + E_f(T_r(d))\right)$ where $C^\dagger > 0$ is a universal constant.*

这些界将在后文证明。现在我们来说明它们为何能让我们界定第二类

错误。由（30）以及引理4 和5，我们得到

$$
\begin{aligned}
P_f\left\{T_r(d) \leq CK_1 n\tau(p,s,n)^2\right\} &\leq \frac{\mathrm{Var}_f(T_r(d))}{\left(E_f(T_r(d)) - CK_1 n\tau(p,s,n)^2\right)^2} \\
&\leq \frac{C^\dagger s^2 r^4 + C^\dagger E_f(T_r(d))}{\left(E_f(T_r(d)) - CK_1 n\tau(p,s,n)^2\right)^2} \\
&\leq \frac{C^\dagger \lceil D \rceil K_2^4 s^2 \nu_{\mathcal{H}} \log\left(1+\frac{p}{s^2}\right)}{C^2 K_1^2 n^2 \tau(p,s,n)^4} + \frac{C^\dagger E_f(T_r(d))}{\left(E_f(T_r(d)) - CK_1 n\tau(p,s,n)^2\right)^2} \\
&= \frac{C^\dagger \lceil D \rceil K_2^4}{C^2 K_1^2} + \frac{C^\dagger E_f(T_r(d))}{\left(E_f(T_r(d)) - CK_1 n\tau(p,s,n)^2\right)^2} \\
&\leq \frac{C^\dagger \lceil D \rceil K_2^4}{C^2 K_1^2} + \frac{C^\dagger E_f(T_r(d))}{\frac{1}{4}\left(E_f(T_r(d))\right)^2} \\
&\leq \frac{C^\dagger \lceil D \rceil K_2^4}{C^2 K_1^2} + \frac{4C^\dagger}{E_f(T_r(d))} \\
&\leq \frac{C^\dagger \lceil D \rceil K_2^4}{C^2 K_1^2} + \frac{4C^\dagger}{2CK_1 n\tau(p,s,n)^2} \\
&\leq \frac{C^\dagger \lceil D \rceil K_2^4}{C^2 K_1^2} + \frac{2C^\dagger}{CK_1\sqrt{\log 2}}
\end{aligned}
$$

只要我们选取 C_η 大于引理所要求的某个充分大的通用常数 4 和5即可。有了这个界，并且由于 $C \geq C_\eta$，现在我们可以选取 C_η 充分大，且只依赖于 η，从而得到

$$
\sup_{\substack{f\in\mathcal{F}_s,\\ ||f||_2 \geq C\tau(p,s,n)}} P_f\left\{T_r(d) \leq CK_1 n\tau(p,s,n)\right\} \leq \frac{\eta}{2}.
$$

总之，我们可以选取 C_η 仅依赖于 η 且充分大，同时满足条件 $C_\eta \geq \log\left(\frac{2}{\eta}\right) \vee 1$ 以及引理 4、5 中的条件，以确保检验风险有界于 η，即

$$
P_0\left\{T_r(d) > CK_1 n\tau(p,s,n)^2\right\} + \sup_{\substack{f\in\mathcal{F}_s,\\ ||f||_2 \geq C\tau(p,s,n)}} P_f\left\{T_r(d) \leq CK_1 n\tau(p,s,n)^2\right\} \leq \eta,
$$

正如所需。 □

剩下的工作是证明引理 4 和 5。回想一下，我们是在命题2的证明所设定的环境下进行讨论。

Proof of Lemma 4. 为证明期望的下界，首先回顾

$$
E_f(T_r(d)) = \sum_{j=1}^{p} E_f\left(\left(E_j(d) - \alpha_r(d)\right)_{\mathbb{I}\{E_j(d)\geq d+r^2\}}\right).
$$

在 P_f 下，我们有 $E_j(d) \sim \chi^2_d(m_j^2)$，其中 $m_j^2 = n\sum_{k\leq d}\theta_{k,j}^2$。这里，集合 $\{\theta_{k,j}\}$ 表示 f 的基系数。直观而言，期望可能较小有两个原因。首先，我们进行了阈值化，因此直观上会去除均值较小但仍非零的那些坐标。此外，我们在 d 处进行截断，不考虑高阶基系数；这也会造成信号损失。

我们先关注阈值化带来的影响。令 $\widetilde{C}$ 表示引理中的通用常数

13. 应用引理 13 可得

$$
\begin{aligned}
E_f(T_r(d)) &= \sum_{j=1}^{p} E_f\left((E_j(d)-\alpha_r(d))_{\mathbb{I}\{E_j(d)\geq d+r^2\}}\right) \\
&= \sum_{j\in S_f} E_f\left((E_j(d)-\alpha_r(d))_{\mathbb{I}\{E_j(d)\geq d+r^2\}}\right) \\
&\geq \sum_{j\in S_f: m_j^2\geq \widetilde{C}r^2} \frac{m_j^2}{2} \\
&= \sum_{j\in S_f} \frac{m_j^2}{2} - \sum_{j\in S_f: m_j^2<\widetilde{C}r^2} \frac{m_j^2}{2} \\
&\geq \left(\sum_{j\in S_f} \frac{m_j^2}{2}\right) - \frac{\widetilde{C}sr^2}{2}
\end{aligned}
$$

其中， $S_f \subset [p]$ 表示活跃变量的子集 j ，满足 $\Theta_j \neq 0$。注意，这样的 S_f 存在，且 $|S_f| \leq s$ ，因为 $f \in \mathcal{F}_s$。至此，我们已界定阈值化损失的信号量。现在来考察截断带来的影响。考虑

$$
\begin{aligned}
||f||_2^2 &\leq \sum_{j\in S_f}\sum_{k=1}^{\infty}\theta_{k,j}^2 \\
&= \sum_{j\in S_f}\left(\sum_{k\leq d}\theta_{k,j}^2 + \sum_{k>d}\theta_{k,j}^2\right) \\
&\leq \sum_{j\in S_f}\left(\sum_{k\leq d}\theta_{k,j}^2 + \mu_{d+1}\sum_{k>d}\frac{\theta_{k,j}^2}{\mu_k}\right) \\
&\leq \left(\sum_{j\in S_f}\sum_{k\leq d}\theta_{k,j}^2\right) + s\mu_{d+1}.
\end{aligned}
$$

这里，我们使用了 $\sum_{k=1}^{\infty}\mu_k^{-1}\theta_{k\,j}^2 \leq 1$ j $\sum_{j\in S_f}\sum_{k\leq d}\theta_{k\,j}^2 \geq$ ，对所有情况都成立。因此，我们证明了 $||f||_2^2 - s\mu_{d+1}$ ，从而量化了截断带来的损失。

现在，我们可以将这两部分结合起来。考虑 $||f||_2^2 \geq C^2\tau(p,s,n)^2$。因此， e-

因此，

$$
\sum_{j\in S_f}\frac{m_j^2}{2} \geq \frac{C^2n\tau(p,s,n)^2 - ns\mu_{d+1}}{2} \geq \frac{C^2n\tau(p,s,n)^2 - ns\mu_{\nu_{\mathcal{H}}+1}}{2}.
$$

我们利用了核的特征值按降序排列这一性质，即 $d \geq \nu_{\mathcal{H}}$ 意味着 $\mu_{\nu_{\mathcal{H}}+1} \geq \mu_{d+1}$。此外，根据 $\nu_{\mathcal{H}}$ 和 $\tau(p,s,n)^2$, 的定义，我们有

$$
\mu_{\nu_{\mathcal{H}}+1} \leq \mu_{\nu_{\mathcal{H}}} \leq \frac{\sqrt{\nu_{\mathcal{H}}\log\left(1+\frac{p}{s^2}\right)}}{n} = \frac{\tau(p,s,n)^2}{s}.
$$

由此可知， $\sum_{j\in S_f}\frac{m_j^2}{2} \geq \left(\frac{C^2-1}{2}\right)n\tau(p,s,n)^2$。综上，我们已经证明

n

$$
\begin{aligned}
E_f(T_r(d)) &\geq \left(\sum_{j\in S_f}\frac{m_j^2}{2}\right) - \frac{\widetilde{C}sr^2}{2} \\
&\geq \left(\sum_{j\in S_f}\frac{m_j^2}{2}\right) - \frac{\widetilde{C}K_2^2\sqrt{\lceil D\rceil}n\tau(p,s,n)^2}{2} \\
&\geq \left(\sum_{j\in S_f}\frac{m_j^2}{2}\right)\left(1 - \frac{\widetilde{C}K_2^2\sqrt{\lceil D\rceil}}{C^2-1}\right). \qquad (31)
\end{aligned}
$$

这里我们使用了 $d \le \lceil D \rceil \nu_{\mathcal{H}}$。我们还可以得出

$$E_f(T_r(d)) \ge \left(\frac{C^2 - 1 - K_2^2 \widetilde{C} \sqrt{\lceil D \rceil}}{2} \right) n\tau(p,s,n)^2.$$

鉴于上述界且由于 $C > C_\eta$，只需选取足够大的 C_η，使其满足 $C_\eta^2 - 4K_1 C_\eta \ge 1 + K_2^2 \widetilde{C}$，从而确保 $E_f(T_r(d)) \ge 2CK_1 n\tau(p,s,n)^2$。引理的证明至此完成。*Proof of Lemma 5.* 为界定 $T_r(d)$ 的方差，回忆 $\sqrt{\log\left(1+\frac{p}{s^2}\right)} \le K_3\sqrt{d}$，因此 $1 \le K_2^2\sqrt{\log\left(1+\frac{p}{s^2}\right)} \le K_2^2 K_3 \sqrt{d}$ ~~$\le L^* \sqrt{d}$~~。于是，我们可以应用引理 14~~。根据引理~~ 14，我们有

$$\begin{aligned}
\mathrm{Var}_f\left(T_r(d)\right) &= \sum_{j=1}^{p} \mathrm{Var}_f\left((E_j(d) - \alpha_r(d))_{\mathbb{I}\{E_j(d) \ge d + r^2\}} \right) \\
&\le C^\dagger p r^4 \exp\left(-\frac{c^{**} r^4}{d} \right) + C^\dagger s r^4 + C^\dagger \sum_{j \in S_f : m_j^2 > 4r^2} m_j^2 \\
&\le C^\dagger p r^4 \exp\left(-\frac{c^{**} r^4}{d} \right) + C^\dagger s r^4 + C^\dagger \sum_{j \in S_f} m_j^2
\end{aligned}$$

其中 $C^\dagger$ 是一个正的通用常数，其取值可能因具体情形而异。回忆一下，c^{**}在命题证明开头定义为2。由于 $r^4 = K_2^4 d \log\left(1+\frac{p}{s^2}\right)$ 且 $c^{**} K_2^4 \ge 1$，我们有

$$p r^4 \exp\left(-\frac{c^{**} r^4}{d} \right) \le p r^4 \exp\left(-\log\left(1 + \frac{p}{s^2}\right) \right) \le p r^4 \cdot \frac{s^2}{s^2 + p} \le s^2 r^4.$$

因此，

$$\mathrm{Var}_f(T_r(d)) \le C^\dagger s^2 r^4 + C^\dagger s r^4 + C^\dagger \sum_{j \in S_f} m_j^2 \le 2C^\dagger s^2 r^4 + C^\dagger \sum_{j \in S_f} m_j^2.$$

取 C_η 大于充分大的通用常数，并注意到 $C \ge C_\eta$，根据（31），即可得到所需结果。

□

Proof of Proposition 3. 我们的证明与命题 2 的证明基本相同，只是我们调用的是关于“尾部”而非“主体部分”的结果。固定 $\eta \in (0,1)$。我们将在证明的最后选取 C_η，因此暂且令 $C > C_\eta$。我们将在证明过程中选取 K_1，但现在选取 K_2。令 $K_2 := \sqrt{\frac{1}{\log 2}} \vee c^{-1/2} \vee \left(cK_3^2\right)^{-1/4}$，其中 $c = c^* \wedge c^{**}$，且 c^* 和 c^{**} 分别是引理 11 和~~10 中~~指数项的通用常数。

我们首先界定第一类错误。由于 $\log\left(1+\frac{p}{s^2}\right) \ge K_3^2 d$，我们有 $r^2 \gtrsim d$。由于 $d \ge D$，根据引理 11，对于任意 $x > 0$,

$$P_0\left\{ T_r(d) > C^* \left(\sqrt{x p r^4 e^{-c^* r^2}} + x \right) \right\} \le e^{-x}$$

其中，C^* 是一个普适正常数。取 $x = C$，并注意到 $C > 1$，在我们已选取$C_\eta \ge 1$的前提下，可得

$$\begin{aligned}
C^* \left(\sqrt{x p r^4 e^{-c^* r^2}} + x \right) &\le C^* C \left(K_2^2 \log\left(1+\frac{p}{s^2}\right) \sqrt{p e^{-c^* K_2^2 \log\left(1+\frac{p}{s^2}\right)}} + 1 \right) \\
&\le C^* C \left(K_2^2 \log\left(1+\frac{p}{s^2}\right) \sqrt{p \cdot \frac{s^2}{s^2+p}} + 1 \right) \\
&\le 2C^* C K_2^2 s \log\left(1+\frac{p}{s^2}\right) \\
&= C K_1 n \tau(p,s,n)^2
\end{aligned}$$

其中我们用到了 $c^* K_2^2 \ge 1$，并设定 $K_1 := 2C^* K_2^2$。因此，采用这些选择

K_1、K_2 和 x，我们有

$$P_0\left\{T_r(d) > CK_1 n\tau(p,s,n)^2\right\} \le e^{-x} = e^{-C} \le e^{-C_\eta} \le \frac{\eta}{2}$$

只要我们选择 $C_\eta \ge \log\left(\frac{2}{\eta}\right) \vee 1$。

现在我们考察第二类错误。为了给第二类错误建立界，我们将使用切比雪夫不等式。具体而言，考虑对于任意 $f \in \mathcal{F}_s$ 满足 $||f||_2 \ge C\tau(p,s,n)$ 的情况，我们有

$$\begin{aligned} P_f\left\{T_r(d) \le CK_1 n\tau(p,s,n)^2\right\} &= P_f\left\{E_f(T_r(d)) - CK_1 n\tau(p,s,n)^2 \le E_f(T_r(d)) - T_r(d)\right\} \\ &\le \frac{\mathrm{Var}_f\left(T_r(d)\right)}{\left(E_f(T_r(d)) - CK_1 n\tau(p,s,n)^2\right)^2} \end{aligned} \tag{32}$$

只要 $E_f(T_r(d)) \ge CK_1 n\tau(p,\ s,\ n)^2$。为了确保切比雪夫不等式的应用有效，并给第二类错误建立界，我们需要 $T_r(d$ 的期望下界。为了给第二类错误建立界，我们还需要 $T_r(d$ 的方差上界。以下引理给出了所需的界；它们与引理 4 和 5 类似，但现在处于尾部区域。

引理 6. *If C_η is larger than some sufficiently larger universal constant, then $E_f(T_r(d)) \ge 2CK_1 n\tau(p,s,n)^2$.*

引理 7. *If C_η is larger than some sufficiently large universal constant, then $\mathrm{Var}_f(T_r(d)) \le C^\dagger(s^2 r^4 + E_f(T_r(d)))$ where $C^\dagger > 0$ is a universal constant.*

有了这些界，基本上可以重复命题 2证明中的论证，从而得出

$$P_0\left\{T_r(d) > CK_1 n\tau(p,s,n)^2\right\} + \sup_{\substack{f\in\mathcal{F}_s,\\ ||f||_2 \ge C\tau(p,s,n)}} P_f\left\{T_r(d) \le CK_1 n\tau(p,s,n)^2\right\} \le \eta$$

在 $C_\eta \ge \log\left(\frac{2}{\eta}\right) \vee 1$ 且 C_η 足够大以满足引理 6 和 7时成立。为简洁起见，省略细节。

□

剩下的工作是证明引理 6 和 7。

Proof of Lemma 6. 证明风格与引理 4的证明类似，但此处要用到尾部区域的结果。令 $\widetilde{C}$ 表示引理 9中的通用常数，应用引理 9，并仿照引理 4 的证明进行推导，可得

$$E_f\left(T_r(d)\right) \ge \left(\sum_{j\in S_f} \frac{m_j^2}{2}\right) - \frac{\widetilde{C} s r^2}{2}$$

其中 $S_f \subset [p]$ 表示活跃变量的子集 j 使得 $\Theta_j \ne 0$。注意，这样的 S_f 存在，且$|S_f| \le s$ 因为 $f \in \mathcal{F}_s$。进一步仿照引理 4 的证明进行推导，并使用 $\sqrt{\log\left(1+\frac{p}{s^2}\right)} \ge K_3\sqrt{d}$，可得

$$\sum_{j\in S_f} \frac{m_j^2}{2} \ge \left(\frac{C^2 - \frac{1}{K_3}}{2}\right) n\tau(p,s,n)^2. \tag{33}$$

总之，我们已经证明

$$\begin{aligned} E_f\left(T_r(d)\right) &\geq \left(\sum_{j\in S_f}\frac{m_j^2}{2}\right) - \frac{\widetilde{C}sr^2}{2} \\ &\geq \left(\sum_{j\in S_f}\frac{m_j^2}{2}\right) - \frac{\widetilde{C}K_2^2 n\tau(p,s,n)^2}{2} \\ &\geq \left(\sum_{j\in S_f}\frac{m_j^2}{2}\right)\left(1-\frac{\widetilde{C}K_2^2}{C^2-\frac{1}{K_3}}\right). \end{aligned} \tag{34}$$

注意，我们还可以得出结论

$$E_f\left(T_r(d)\right) \geq \left(\frac{C^2-\frac{1}{K_3}-K_2^2\widetilde{C}}{2}\right) n\tau(p,s,n)^2.$$

由于 $C > C_\eta$，只需取足够大的 C_η，使其满足 $C_\eta^2 - 4K_1C_\eta \geq \frac{1}{K_3} + K_2^2\widetilde{C}$，以确保 $E_f(T_r(d))$ $\geq 2CK_1 n\tau(p,s,n)^2$。引理得证。 □

Proof of Lemma 7. 回顾一下，$\log\left(1+\frac{p}{s^2}\right) \geq K_3^2 d$。根据 r^2 的定义，我们有 $r^2 \geq K_2^2K_3^2 d$。换言之，$r^2 \gtrsim d$ 因此我们可以应用引理 10。根据引理 10，我们有

$$\begin{aligned} \mathrm{Var}_f\left(T_r(d)\right) &= \sum_{j=1}^{p}\mathrm{Var}_f\left(\left(E_j(d)-\alpha_r(d)\right)_{\mathbb{I}\{E_j(d)\geq d+r^2\}}\right) \\ &\leq C^\dagger pr^4\exp\left(-c^{**}\min\left(\frac{r^4}{d},r^2\right)\right) + C^\dagger sr^4 + C^\dagger\sum_{j\in S_f:m_j^2>4r^2} m_j^2 \\ &\leq C^\dagger pr^4\exp\left(-c^{**}\min\left(\frac{r^4}{d},r^2\right)\right) + C^\dagger sr^4 + C^\dagger\sum_{j\in S_f} m_j^2. \end{aligned}$$

其中 $C^\dagger$ 是一个正的通用常数。回顾一下，在命题 3的证明开始时，我们定义了 c^{**}。由于 $r^2 \geq K_2^2K_3^2 d$，我们有 $\frac{r^4}{d} \geq r^2K_2^2K_3^2$。因此，

$$\exp\left(-c^{**}\min\left(\frac{r^4}{d},r^2\right)\right) \leq \exp\left(-c^{**}\left(K_2^2K_3^2\wedge 1\right)r^2\right) \leq \exp\left(-\log\left(1+\frac{p}{s^2}\right)\right) \leq \frac{s^2}{s^2+p}$$

其中我们利用了根据 K_2 的定义，$c^{**}\left(K_2^2K_3^2\wedge 1\right)K_2^2 \geq 1$。因此，

$$\mathrm{Var}_f\left(T_r(d)\right) \leq 2C^\dagger s^2r^4 + C^\dagger\sum_{j\in S_f} m_j^2.$$

取 C_η 大于一个足够大的通用常数，并应用（34），即可得到所需结果。

□

7.1.2 稠密情形

Proof of Proposition 4. 为简化记号，令 $L = \frac{2}{\sqrt{\eta}}$。定义 $C_\eta := \left(\sqrt{2\sqrt{2}L+1}\right)\vee\left(\sqrt{1+\frac{4}{\sqrt{\eta}}+L}\right)\vee$ $\left(\sqrt{\frac{64\sqrt{2}}{\eta}+1}\right)$。令 $C > C_\eta$。我们首先界定第一类错误。注意，在 P_0 下，我们有 $T \sim \chi^2_{p\nu_{\mathcal{H}}}$。注意到 E_0 $(T) = p\nu_{\mathcal{H}}$，且 $\mathrm{Var}_0(T) = 2p\nu_{\mathcal{H}}$。因此，根据切比雪夫不等式，我们有

$$P_0\left\{T > p\nu_{\mathcal{H}} + L\sqrt{p\nu_{\mathcal{H}}}\right\} \leq \frac{\mathrm{Var}_0(T)}{L^2p\nu_{\mathcal{H}}} = \frac{2}{L^2} \leq \frac{\eta}{2}.$$

现在我们考察第二类错误。在 P_f 下，我们有 $T \sim \chi^2_{p\nu_{\mathcal{H}}}\left(n\sum_{j=1}^{p}\sum_{k\leq\nu_{\mathcal{H}}}\theta_{k,j}^2\right)$ 其中 $\{\theta_{k,j}\}$ 表示与 f 相关联的基系数。注意 $E_f(T) = p\nu_{\mathcal{H}} + n\sum_{j=1}^{p}\sum_{k\leq\nu_H}\theta_{k,\,j}^2$ 且 $\mathrm{Var}f(T)$ $= 2p\nu_{\mathcal{H}} + 4n\sum_{j=1}^{p}\sum_{k\leq\nu_{\mathcal{H}}}\theta_{k,\,j}^2$。为了继续论证，我们需要得到信号强度的一个下界估计。令 S 表示满足 Θ_j 非零的 j 的集合，

考虑对于 $||f||_2^2 \geq C^2\tau(p,s,n)^2$，我们有

$$\begin{aligned}
C^2\tau(p,s,n)^2 &\leq ||f||_2^2 \\
&= \sum_{j=1}^{p}\sum_{k=1}^{\infty}\theta_{k,j}^2 \\
&\leq \sum_{j\in S}\left(\sum_{k\leq\nu_{\mathcal{H}}}\theta_{k,j}^2 + \sum_{k>\nu_{\mathcal{H}}}\theta_{k,j}^2\right) \\
&\leq \sum_{j\in S}\left(\sum_{k\leq\nu_{\mathcal{H}}}\theta_{k,j}^2 + \mu_{\nu_{\mathcal{H}}}\sum_{k>\nu_{\mathcal{H}}}\frac{\theta_{k,j}^2}{\mu_k}\right) \\
&\leq \left(\sum_{j\in S}\sum_{k\leq\nu_{\mathcal{H}}}\theta_{k,j}^2\right) + s\mu_{\nu_{\mathcal{H}}}.
\end{aligned}$$

在最后一行中，我们对所有 j 使用了 $\sum_{k=1}^{\infty}\frac{\theta_{k,j}^2}{\mu_k} \leq 1$。因此，根据 $\nu_{\mathcal{H}}$ 的定义，我们有

$$n\sum_{j\in S}\sum_{k\leq\nu_{\mathcal{H}}}\theta_{k,j}^2 \geq nC^2\tau(p,s,n)^2 - ns\mu_{\nu_{\mathcal{H}}} \geq \left(C_\eta^2 - 1\right)\sqrt{\nu_{\mathcal{H}}p}$$

其中我们使用了 $ns\mu_{\nu_{\mathcal{H}}} \leq s\sqrt{\nu_{\mathcal{H}}}\log(1+\frac{p}{s^2}) \leq \sqrt{p\nu_{\mathcal{H}}}$。

现在我们继续界定第二类错误。根据切比雪夫不等式，我们有

$$\begin{aligned}
&\sup_{\substack{f\in\mathcal{F}_s,\\ ||f||_2\geq C_\eta\tau(p,s,n)}} P_f\left\{T \leq p\nu_{\mathcal{H}} + L\sqrt{p\nu_{\mathcal{H}}}\right\} \\
=& \sup_{\substack{f\in\mathcal{F}_s,\\ ||f||_2\geq C_\eta\tau(p,s,n)}} P_f\left\{E_f(T) - p\nu_{\mathcal{H}} - L\sqrt{p\nu_{\mathcal{H}}} \leq E_f(T) - T\right\} \\
\leq& \sup_{\substack{f\in\mathcal{F}_s,\\ ||f||_2\geq C_\eta\tau(p,s,n)}} \frac{\mathrm{Var}_f\left(T\right)}{\left(E_f(T) - p\nu_{\mathcal{H}} - L\sqrt{p\nu_{\mathcal{H}}}\right)^2} \\
\leq& \sup_{\substack{f\in\mathcal{F}_s,\\ ||f||_2\geq C_\eta\tau(p,s,n)}} \frac{2p\nu_{\mathcal{H}} + 4n\sum_{j=1}^{p}\sum_{k\leq\nu_{\mathcal{H}}}\theta_{k,j}^2}{\left(n\sum_{j=1}^{p}\sum_{k\leq\nu_{\mathcal{H}}}\theta_{k,j}^2 - L\sqrt{p\nu_{\mathcal{H}}}\right)^2} \\
=& \sup_{\substack{f\in\mathcal{F}_s,\\ ||f||_2\geq C_\eta\tau(p,s,n)}} \frac{2p\nu_{\mathcal{H}}}{\left(C_\eta^2 - 1 - L\right)^2\nu_{\mathcal{H}}p} + \frac{4n\sum_{j=1}^{p}\sum_{k\leq\nu_{\mathcal{H}}}\theta_{k,j}^2}{\left(n\sum_{j=1}^{p}\sum_{k\leq\nu_{\mathcal{H}}}\theta_{k,j}^2 - L\sqrt{p\nu_{\mathcal{H}}}\right)^2} \\
\leq& \sup_{\substack{f\in\mathcal{F}_s,\\ ||f||_2\geq C_\eta\tau(p,s,n)}} \frac{2}{\left(C_\eta^2 - 1 - L\right)^2} + \frac{4n\sum_{j=1}^{p}\sum_{k\leq\nu_{\mathcal{H}}}\theta_{k,j}^2}{\left(\frac{1}{2}n\sum_{j=1}^{p}\sum_{k\leq\nu_{\mathcal{H}}}\theta_{k,j}^2\right)^2} \\
=& \sup_{\substack{f\in\mathcal{F}_s,\\ ||f||_2\geq C_\eta\tau(p,s,n)}} \frac{2}{\left(C_\eta^2 - 1 - L\right)^2} + \frac{16}{n\sum_{j=1}^{p}\sum_{k\leq\nu_{\mathcal{H}}}\theta_{k,j}^2} \\
\leq& \frac{2}{\left(C_\eta^2 - 1 - L\right)^2} + \frac{16}{(C_\eta^2 - 1)\sqrt{\nu_{H}p}} \\
\leq& \frac{2}{\left(C_\eta^2 - 1 - L\right)^2} + \frac{16}{C_\eta^2 - 1} \\
\leq& \frac{\eta}{4} + \frac{\eta}{4} \\
\leq& \frac{\eta}{2}.
\end{aligned}$$

因此，第一类错误与第二类错误之和被 η 所界定，正如所要求的。 □

7.2 极小极大下界

Proof of Proposition 1. 我们将分析分为两种情况。

情形 1：假设 $s < \sqrt{p}$。我们将在 $\mathscr{T}_s$ 上构造先验分布 π，并使用 Le Cam 二点

方法来给出下界。定义 $c_\eta := 1 \wedge \sqrt{\kappa} \wedge \sqrt{\kappa \log(1+4\eta^2)}$。令 $0 < c < c_\eta$。令 π 为这样的先验：通过均匀抽取大小为 s 的 $S \subset [p]$，构造一次抽取结果 $\Theta \sim \pi$，并设置

$$\Theta_j = \begin{cases} ce_1 & \text{if } j \in S, \\ 0 & \text{otherwise} \end{cases}$$

其中，$e_1 \in \mathbb{R}^{\mathbb{N}}$ 由 $e_1 = (1,0,0,\ldots)$给出。注意到，$\Theta \sim \pi$ 蕴含 $\|\Theta\|_F^2 = c^2 s$ 和 $\Theta \in \mathscr{T}_s$。根据 Neyman-Pearson 引理以及不等式 $1 - d_{TV}(Q,P) \geq 1 - \sqrt{\chi^2(Q\|P)/2}$，我们有

$$\inf_{\varphi} \left\{ P_0\{\varphi \neq 0\} + \sup_{\substack{\Theta \in \mathscr{T}_s, \\ \|\Theta\|_F \geq c\sqrt{s}}} P_\Theta\{\varphi \neq 1\} \right\} \geq 1 - \frac{1}{2}\sqrt{\chi^2(P_\pi \| P_0)} \tag{35}$$

其中，P_π 表示由 $\int P_\Theta \, \pi(d\Theta)$ 诱导的混合分布 π。根据 Ingster-Suslina 方法（命题5）和推论3，我们有

$$\begin{aligned} \chi^2(P_\pi \| P_0) &= E\left(\exp\left(n\langle \Theta, \widetilde{\Theta}\rangle_F\right)\right) - 1 \\ &= E\left(\exp\left(c^2 n |S \cap \widetilde{S}|\right)\right) - 1 \\ &\leq \left(1 - \frac{s}{p} + \frac{s}{p} e^{c^2 n}\right)^s - 1. \end{aligned}$$

其中，Θ、$\widetilde{\Theta} \overset{iid}{\sim} \pi$ 和 S、$\widetilde{S}$ 分别是相应的随机集合。现在，利用 $\log\left(1 + \frac{p}{s^2}\right) \geq \kappa n$，我们有

$$\begin{aligned} &\leq \left(1 - \frac{s}{p} + \frac{s}{p} e^{c^2 \kappa^{-1} \log\left(1 + \frac{p}{s^2}\right)}\right)^s - 1 \\ &= \left(1 - \frac{s}{p} + \frac{s}{p}\left(1 + \frac{p}{s^2}\right)^{c^2\kappa^{-1}}\right)^s - 1 \\ &\leq \left(1 - \frac{s}{p} + \frac{s}{p} + \frac{c^2\kappa^{-1}}{s}\right)^s - 1 \\ &\leq \exp\left(\frac{c^2}{\kappa}\right) - 1 \\ &\leq 4\eta^2 \end{aligned}$$

我们还利用了 $c^2\kappa^{-1} < c_\eta^2 \kappa^{-1} \leq 1$ 以及不等式 $(1+x)^y \leq 1 + xy$ 对于 $x > 0$ 和 $y \in (0,1)$成立。代入（35）即可得到所需结果。

情形 2：假设 $s \geq \sqrt{p}$。我们可以重复情形 1 中的分析，只需将每处 s 替换为 $\lceil \sqrt{p} \rceil$。

□

Proof of Theorem 1. 我们将构造定义在 $\mathscr{T}_s$ 上的先验分布 π，并使用 Le Cam 二点方法给出一个下界。令 $c_\eta := 1 \wedge 2^{1/4} \wedge \left(\log\left(1+4\eta^2\right)\right)^{1/4}$。设 $0 < c < c_\eta$。我们将分析分为两种情形。

情形 1：假设我们处于满足 $^2 = s\Gamma_{\mathcal{H}}$ 的情形，其中 $\psi(p,s,n)$。令 $\rho := \sqrt{\frac{\Gamma_{\mathcal{H}}}{\nu_{\mathcal{H}}}}$。设 π 为先验分布，其中一次抽取得到的 $\Theta \sim \pi$ 是通过从大小为 s 的 $S \subset [p]$ 中均匀抽取得到的，并抽取

$$\theta_{i,k} \sim \begin{cases} \text{Uniform}\{-c\rho, c\rho\} & \text{if } i \in S \text{ and } k < \nu_{\mathcal{H}}, \\ \delta_0 & \text{otherwise} \end{cases}$$

其中，δ_0 表示将全部概率质量集中在零点的概率测度。注意，$\Theta \sim \pi$ 意味着 $\|\Theta\|_F^2 = c^2\rho^2 s(\nu_{\mathcal{H}} - 1)$ $= c^2 s\Gamma_{\mathcal{H}} \cdot \frac{\nu_{\mathcal{H}} - 1}{\nu_{\mathcal{H}}} \geq \frac{c^2}{2} s\Gamma_{\mathcal{H}}$。这里我们利用了 $\nu_{\mathcal{H}} \geq 2$，因为 $\log\left(1 + \frac{p}{s^2}\right) \leq \frac{n}{2}$。此外，考虑到对于 $i \in S$，我们有

$$\sum_{\ell=1}^{\infty} \frac{\theta_{i,\ell}^2}{\mu_\ell} = c^2\rho^2 \sum_{\ell < \nu_{\mathcal{H}}} \frac{1}{\mu_\ell} = c^2 \frac{1}{\nu_{\mathcal{H}}} \sum_{\ell < \nu_{\mathcal{H}}} \frac{\Gamma_{\mathcal{H}}}{\mu_\ell} \leq c^2 \frac{1}{\nu_{\mathcal{H}}} \sum_{\ell < \nu_{\mathcal{H}}} \frac{\mu_{\nu_{\mathcal{H}}-1}}{\mu_\ell} \leq c^2 \leq c_\eta^2 \leq 1.$$

这里我们利用了特征值的大小顺序 $\mu_1 \geq \mu_2 \geq \ldots \geq 0$ 和 $\Gamma_{\mathcal{H}} \leq \mu_{\nu_{\mathcal{H}}-1}$ 根据引理25。当然，对于 $i \notin S$ 我们有 $\Theta_i = 0$。因此我们有 $\Theta \in \mathscr{T}_s$。然后根据 Neyman-Pearson 引理以及不等式 $1 - d_{TV}(Q, P) \geq 1 - \sqrt{\chi^2(Q||P)}/2$ 可得

$$\inf_{\varphi} \left\{ P_0\{\varphi \neq 0\} + \sup_{\substack{\Theta \in \mathscr{T}_s, \\ ||\Theta||_F \geq c\psi(p,s,n)}} P_\Theta\{\varphi \neq 1\} \right\} \geq 1 - \frac{1}{2}\sqrt{\chi^2(P_\pi||P_0)}. \tag{36}$$

在以下计算中，设 Θ，$\Theta' \overset{iid}{\sim} \pi$ 且 S，S' 为相应的随机集合。另外，设 $\{r_{i\ell}, r'_{i\ell}\}_{1\leq i\leq p,\ \ell\in\mathbb{N}}$ 表示一组独立同分布的 Rademacher（1/2）随机变量。根据 Ingster-Suslina 方法（命题 5），利用 $\{S, S'\}$ 与 $\{r_{i\ell}, r'_{i\ell}\}_{1\leq i\leq p\text{的独立性、}\ell\in\mathbb{N}}$ 以及推论 3，可得

$$\begin{aligned}
\chi^2(P_\pi||P_0) &= E\left(\exp\left(n\langle\Theta,\Theta'\rangle_F\right)\right) - 1 \\
&= E\left(\exp\left(nc^2\rho^2 \sum_{i\in S\cap S'}\sum_{\ell\leq\nu_{\mathcal{H}}} r_{i\ell}r'_{i\ell}\right)\right) - 1 \\
&= E\left(\prod_{i\in S\cap S'} E\left(\exp\left(nc^2\rho^2 \sum_{\ell\leq\nu_{\mathcal{H}}} r_{i\ell}r'_{i\ell}\right)\right)\right) - 1 \\
&= E\left(\cosh\left(nc^2\rho^2\right)^{\nu_{\mathcal{H}}|S\cap S'|}\right) - 1 \\
&\leq E\left(\exp\left(\frac{c^4n^2\rho^4\nu_{\mathcal{H}}}{2}|S\cap S'|\right)\right) - 1 \\
&\leq \left(1 - \frac{s}{p} + \frac{s}{p}e^{\frac{c^4n^2\rho^4\nu_{\mathcal{H}}}{2}}\right)^s - 1.
\end{aligned}$$

根据引理 1 可知，$c^4n^2\rho^4\nu_{\mathcal{H}} = \frac{c^4n^2\Gamma_{\mathcal{H}}^2}{\nu_{\mathcal{H}}} \leq c^4\log\left(1+\frac{p}{s^2}\right)$。因此，

$$\left(1 - \frac{s}{p} + \frac{s}{p}e^{\frac{c^4n^2\rho^4\nu_{\mathcal{H}}}{2}}\right)^s - 1 \leq \left(1 - \frac{s}{p} + \frac{s}{p}e^{\frac{c^4}{2}\log\left(1+\frac{p}{s^2}\right)}\right)^s - 1 \leq \left(1 + \frac{c^4}{2s}\right)^s - 1 \leq e^{c_\eta^4/2} - 1 = 4\eta^2.$$

我们使用了 $\frac{c^4}{2} < \frac{c_\eta^4}{2} \leq 1$ 以及不等式 $(1+x)^y \leq 1 + xy$（适用于 $x > 0$ 和 $y \in (0, 1)$）。使用 $\chi^2(P_\pi||P_0) \leq 4\eta^2$，结合式（36），即可得到所需结果。

情形 2：假设 $s < \sqrt{p}$ 和 $\psi(p,s,n)^2 = \frac{s}{n}\log\left(1+\frac{p}{s^2}\right)$。令 $\rho = 1$。设 π 为这样一个先验：通过均匀抽取大小为 s 的 $S \subset [p]$ 并抽取，得到一个抽样 $\Theta \sim \pi$

$$\theta_{i,k} \sim \begin{cases} \text{Uniform}\{-c\rho, c\rho\} & \text{if } i \in S \text{ and } k = 1, \\ \delta_0 & \text{otherwise.} \end{cases}$$

所需结果可通过采用与定理1 的证明类似的论证加以证明（另见 [10]）。为简洁起见，省略具体细节。

□

7.3 自适应下界

Proof of Theorem 2. 回顾（15）中的 $\mathscr{A}_{\mathcal{H}}$ 和（20）中的 $\widetilde{\mathscr{V}}_{\mathcal{H}}$ 的定义。根据假设，对于某个通用常数 $L > 0$，我们有 $\mathscr{A}_{\mathcal{H}} \leq L\log(e|\widetilde{\mathscr{V}}_{\mathcal{H}}|)$。为方便记号，记 $\psi = \psi_{\text{adapt}}(p,s,n)$，其中 ψ_{adapt} 由（19）给出。令 $c_\eta := (16L)^{-1/4} \wedge \left(\frac{\eta^2}{32L}\log\frac{1+2\eta^2}{1}\right)^{1/4}$，并令 $0 < c < c_\eta$。现在我们定义一个先验 π。一次抽样 $\Theta \sim \pi$ 按如下方式获得。首先，从均匀分布 $\text{Uniform}(\widetilde{\mathscr{V}}_{\mathcal{H}})$ 中抽取 $v \sim$。令 $s := \min\left\{s' \in [p] : s' \geq \sqrt{p\mathscr{A}_{\mathcal{H}}} \text{ 和 } \frac{v}{2} < \nu_{\mathcal{H}}(s', \mathscr{A}_{\mathcal{H}}) \leq v\right\}$。然后，从大小恰好为 s 的子集中均匀随机抽取一个子集 $S \subset [p]$。令 $d_s := 2^{k_s} \in \widetilde{\mathscr{V}}_{\mathcal{H}}$ 满足 $2^{k_s-1} < \nu_{\mathcal{H}}(s, \mathscr{A}_{\mathcal{H}}) \leq 2^{k_s}$。

独立抽取

$$\theta_{k,j} \sim \begin{cases} \text{Uniform}\left\{-\sqrt{2}c\rho_s, \sqrt{2}c\rho_s\right\} & \text{if } j \in S \text{ and } 1 \le k < \nu_{\mathcal{H}}(s, \mathscr{A}_{\mathcal{H}}), \\ \delta_0 & \text{otherwise.} \end{cases}$$

这里， $\rho_s = \sqrt{\frac{\psi^2/s}{\nu_{\mathcal{H}}(s, \mathscr{A}_{\mathcal{H}})}}$。该描述结束了对 π 的定义。

现在我们必须证明， π 确实支撑在 $\bigcup_s \mathscr{T}_s$ 上。注意，当我们绘制 $\Theta \sim \pi$ 时，相应的稀疏度 s 始终满足 $s \ge \sqrt{p\mathscr{A}_{\mathcal{H}}}$。此外，在给定 s 的条件下，我们有

$$||\Theta||_F^2 = 2c^2 \frac{\psi^2}{\nu_{\mathcal{H}}(s, \mathscr{A}_{\mathcal{H}})}(\nu_{\mathcal{H}}(s, \mathscr{A}_{\mathcal{H}}) - 1) \ge c^2\psi^2.$$

此外，注意对于 $j \in S$，我们有

$$\begin{aligned} \sum_{\ell=1}^{\infty} \frac{\theta_{\ell,j}^2}{\mu_\ell} &\le 2c^2\rho^2 \sum_{\ell < \nu_{\mathcal{H}}(s, \mathscr{A}_{\mathcal{H}})} \frac{1}{\mu_\ell} \\ &= 2c^2 \frac{1}{\nu_{\mathcal{H}}(s, \mathscr{A}_{\mathcal{H}})} \sum_{\ell < \nu_{\mathcal{H}}(s, \mathscr{A}_{\mathcal{H}})} \frac{\psi^2/s}{\mu_\ell} \\ &\le 2c^2 \frac{1}{\nu_{\mathcal{H}}(s, \mathscr{A}_{\mathcal{H}})} \sum_{\ell < \nu_{\mathcal{H}}(s, \mathscr{A}_{\mathcal{H}})} \frac{\mu_{\nu_{\mathcal{H}}(s, \mathscr{A}_{\mathcal{H}})}}{\mu_\ell} \\ &\le 2c^2 \\ &\le 1 \end{aligned}$$

其中，最后一行由 $c < c_\eta$推出。注意，我们使用了特征值的排序，以及$\Gamma_{\mathcal{H}}(s, \mathscr{A}_{\mathcal{H}}) \le \mu_{\nu_{\mathcal{H}}(s, \mathscr{A}_{\mathcal{H}})-1}$，根据引理 26。当然，对于 $j \notin S$，我们有 $\Theta_j = 0$。因此， $\Theta \in \mathscr{T}_s$ 从而 π 得到了正确的支撑。

将 $P_\pi = \int P_\Theta \pi(d\Theta)$ 记作该混合分布，我们有

$$\inf_{\varphi} \left\{ P_0\{\varphi = 1\} + \max_{s \ge \sqrt{p\mathscr{A}_{\mathcal{H}}}} \sup_{\substack{\Theta \in \mathscr{T}_s, \\ ||\Theta||_F \ge c\psi_{\text{adapt}}(p,s,n)}} P_\Theta\{\varphi = 0\} \right\} \ge 1 - \frac{1}{2}\sqrt{\chi^2(P_\pi \,||\, P_0)}. \tag{37}$$

在接下来的计算中，设 $\Theta, \Theta' \overset{iid}{\sim} \pi$。令 $v, v' \in \widetilde{\mathscr{V}}_{\mathcal{H}}$ 为对应的量。令 s和 t 表示对应的稀疏度。将对应的支撑集记为 S 和 T。此外，令$\{r_{i\ell}, \tilde{r}_{i\ell}\}_{1 \le i \le p, \ell \in \mathbb{N}}$ 表示一组独立同分布的 Rademacher $(1/2)$ 随机变量，且与 s, t, S和 T独立。为便于记号，令 $\nu_s = \nu_{\mathcal{H}}(s, \mathscr{A}_{\mathcal{H}})$。类似地，令 $\nu_t = \nu_{\mathcal{H}}(t, \mathscr{A}_{\mathcal{H}})$。根据 Ingster-Suslina 方法（命题 5），我们有

$$\begin{aligned} \chi^2(P_\pi \,||\, P_0) + 1 &= E\left(\exp\left(n\langle\Theta, \Theta'\rangle_F\right)\right) \\ &= E\left(\exp\left(2nc^2\rho_s\rho_t \sum_{i \in S \cap T} \sum_{1 < \ell < \nu_s \wedge \nu_t} r_{i\ell}\tilde{r}_{i\ell}\right)\right) \\ &= E\left(\prod_{i \in S \cap T} \prod_{1 < \ell < \nu_s \wedge \nu_t} \cosh\left(2nc^2\rho_s\rho_t\right)\right) \\ &\le E\left(\exp\left(2n^2c^4\rho_s^2\rho_t^2(\nu_s \wedge \nu_t)|S \cap T|\right)\right) \\ &\le E\left(\exp\left(2n^2c^4\rho_s^2\rho_t^2(d_s \wedge d_t)|S \cap T|\right)\right). \end{aligned}$$

这里，我们使用了不等式 $\cosh(x) \le e^{x^2/2}$ 对于 $x > 0$。注意，根据引理 26

$$\begin{aligned}
2n^2c^4\rho_s^2\rho_t^2(d_s \wedge d_t) &= 2n^2c^4 \cdot \frac{\Gamma_{\mathcal{H}}(s, \mathscr{A}_{\mathcal{H}})}{\nu_s} \frac{\Gamma_{\mathcal{H}}(t, \mathscr{A}_{\mathcal{H}})}{\nu_t} \cdot (d_s \wedge d_t) \\
&\le 2c^4 \sqrt{\log\left(1 + \frac{p\mathscr{A}_{\mathcal{H}}}{s^2}\right) \log\left(1 + \frac{p\mathscr{A}_{\mathcal{H}}}{t^2}\right)} \cdot \frac{d_s \wedge d_t}{\sqrt{\nu_s \nu_t}} \\
&\le 4c^4 \sqrt{\log\left(1 + \frac{p\mathscr{A}_{\mathcal{H}}}{s^2}\right) \log\left(1 + \frac{p\mathscr{A}_{\mathcal{H}}}{t^2}\right)} \cdot \frac{d_s \wedge d_t}{\sqrt{d_s d_t}} \\
&\le 4c^4 \sqrt{\frac{p\mathscr{A}_{\mathcal{H}}}{s^2} \cdot \frac{p\mathscr{A}_{\mathcal{H}}}{t^2}} \cdot \frac{d_s \wedge d_t}{\sqrt{d_s d_t}} \\
&\le 8c^4 \log\left(1 + \frac{p\mathscr{A}_{\mathcal{H}}}{st}\right) \cdot \frac{d_s \wedge d_t}{\sqrt{d_s d_t}}
\end{aligned}$$

我们结合使用了 $s,t \ge \sqrt{p\mathscr{A}_{\mathcal{H}}}$ 以及不等式 $\frac{u}{2} \le \log(1+u) \le u$ for $0 \le u \le 1$。因此，

$$E\left(\exp\left(n\langle \Theta, \Theta' \rangle_F\right)\right) \le E\left(\exp\left(8c^4 \log\left(1 + \frac{p\mathscr{A}_{\mathcal{H}}}{st}\right) \cdot \frac{d_s \wedge d_t}{\sqrt{d_s d_t}} |S \cap T|\right)\right). \tag{38}$$

由于 $8c^4 \frac{d_s \wedge d_t}{\sqrt{d_s d_t}} \le 1$，我们可以使用引理 28 和不等式 $(1+x)^\delta \le 1 + \delta x$ 对于 $x > 0$ 和$\delta \le 1$，得到

$$\begin{aligned}
\chi^2\left(P_\pi \,||\, P_0\right) + 1 &\le E\left(\exp\left(8c^4 \frac{d_s \wedge d_t}{\sqrt{d_s d_t}} \log\left(1 + \frac{p\mathscr{A}_{\mathcal{H}}}{st}\right) |S \cap T|\right)\right) \\
&\le E\left(\left(1 - \frac{s}{p} + \frac{s}{p} \exp\left(8c^4 \log\left(1 + \frac{p\mathscr{A}_{\mathcal{H}}}{st}\right) \cdot \frac{d_s \wedge d_t}{\sqrt{d_s d_t}}\right)\right)^t\right) \\
&= E\left(\left(1 - \frac{s}{p} + \frac{s}{p}\left(1 + \frac{p\mathscr{A}_{\mathcal{H}}}{st}\right)^{8c^4 \frac{d_s \wedge d_t}{\sqrt{d_s d_t}}}\right)^t\right) \\
&\le E\left(\left(1 + \frac{1}{t} \cdot 8c^4 \frac{d_s \wedge d_t}{\sqrt{d_s d_t}} \mathscr{A}_{\mathcal{H}}\right)^t\right) \\
&\le E\left(\exp\left(8c^4 \frac{d_s \wedge d_t}{\sqrt{d_s d_t}} \mathscr{A}_{\mathcal{H}}\right)\right).
\end{aligned}$$

回忆一下，我们记作 $d_s = 2^{k_s}$ 和 $d_t = 2^{k_t}$。此外，回忆 $d_s, d_t \in \widetilde{\mathscr{V}}_{\mathcal{H}}$。现在注意到

$$E\left(\exp\left(8c^4 \frac{d_s \wedge d_t}{\sqrt{d_s d_t}} \mathscr{A}_{\mathcal{H}}\right)\right) = E\left(\exp\left(8c^4 2^{-\frac{|k_s - k_t|}{2}} \mathscr{A}_{\mathcal{H}}\right)\right).$$

回忆一下，我们有 $\mathscr{A}_{\mathcal{H}} \le L\log(e|\widetilde{\mathscr{V}}_{\mathcal{H}}|)$。定义这些集合

$$\begin{aligned}
E_1 &:= \left\{(k_s, k_t) \in \mathbb{N} \times \mathbb{N} : 2^{k_s}, 2^{k_t} \in \widetilde{\mathscr{V}}_{\mathcal{H}} \text{ and } |k_s - k_t| \le \frac{\eta^2}{2e^{8Lc^4}} \log_2(e|\widetilde{\mathscr{V}}_{\mathcal{H}}|)\right\}, \\
E_2 &:= \left\{(k_s, k_t) \in \mathbb{N} \times \mathbb{N} : 2^{k_s}, 2^{k_t} \in \widetilde{\mathscr{V}}_{\mathcal{H}} \text{ and } |k_s - k_t| > \frac{\eta^2}{2e^{8Lc^4}} \log_2(e|\widetilde{\mathscr{V}}_{\mathcal{H}}|)\right\}.
\end{aligned}$$

Then

$$\begin{aligned}
&E\left(\exp\left(8c^4 2^{-\frac{|k_s - k_t|}{2}} \mathscr{A}_{\mathcal{H}}\right)\right) \\
&= E\left(\exp\left(8c^4 2^{-\frac{|k_s - k_t|}{2}} \mathscr{A}_{\mathcal{H}}\right) \mathbb{I}^{\{(k_s,k_t) \in E_1\}}\right) + E\left(\exp\left(8c^4 2^{-\frac{|k_s - k_t|}{2}} \mathscr{A}_{\mathcal{H}}\right) \mathbb{I}^{\{(k_s,k_t) \in E_2\}}\right).
\end{aligned}$$

我们来考察第二项。注意，对于任意 $\ell > 0$，我们都有 $x^{-\ell}\log(x) \le (e\ell)^{-1}$，对所有 $x >$均成立 0.

利用这一点，我们得到

$$
\begin{aligned}
E\left(\exp\left(8c^4 2^{-\frac{|k_s-k_t|}{2}}\mathscr{A}_{\mathcal{H}}\right)\mathbb{I}^{\{(k_s,k_t)\in E_2\}}\right) &\le \exp\left(8Lc^4\left(e|\widetilde{\mathscr{V}}_{\mathcal{H}}|\right)^{-\frac{\eta^2}{4e^{8Lc^4}}}\mathscr{A}_{\mathcal{H}}\right)\\
&\le \exp\left(8Lc^4\left(e|\widetilde{\mathscr{V}}_{\mathcal{H}}|\right)^{-\frac{\eta^2}{4e^{8Lc^4}}}\log(e|\widetilde{\mathscr{V}}_{\mathcal{H}}|)\right)\\
&\le \exp\left(8Lc^4\cdot\frac{4e^{8Lc^4-1}}{\eta^2}\right)\\
&\le \exp\left(\frac{32Lc^4}{\eta^2}\right)\\
&\le 2\eta^2+1.
\end{aligned}
$$

最后一行由 $c^4 < c_\eta^4 \le \frac{1}{32L}\eta^2\log\left(1+2\eta^2\right)$ 得出。注意，我们还使用了 $8Lc^4-1\le 0$ 来得到倒数第二行。现在我们考察 E_1。注意到 $|E_1|\le \frac{\eta^2}{2e^{8Lc^4}}|\widetilde{\mathscr{V}}_{\mathcal{H}}|\log_2(e|\widetilde{\mathscr{V}}_{\mathcal{H}}|)$。因此，

$$
\begin{aligned}
E\left(\exp\left(8c^4 2^{-\frac{|k_s-k_t|}{2}}\mathscr{A}_{\mathcal{H}}\right)\mathbb{I}^{\{(k_s,k_t)\in E_1\}}\right) &= \frac{1}{|\widetilde{\mathscr{V}}_{\mathcal{H}}|^2}\sum_{v,v'\in\widetilde{\mathscr{V}}_{\mathcal{H}}}\exp\left(8c^4 2^{-\frac{|k_s-k_t|}{2}}\mathscr{A}_{\mathcal{H}}\right)\mathbb{I}^{\{(2^{k_s},2^{k_t})\in E_1\}}\\
&\le \frac{|E_1|}{|\widetilde{\mathscr{V}}_{\mathcal{H}}|^2}\exp\left(8c^4\mathscr{A}_{\mathcal{H}}\right)\\
&\le \frac{\frac{\eta^2}{2e^{8Lc^4}}|\widetilde{\mathscr{V}}_{\mathcal{H}}|\log_2(e|\widetilde{\mathscr{V}}_{\mathcal{H}}|)}{|\widetilde{\mathscr{V}}_{\mathcal{H}}|^2}\cdot\exp\left(8Lc^4\log(e|\widetilde{\mathscr{V}}_{\mathcal{H}}|)\right)\\
&\le \frac{\eta^2}{2}\cdot\frac{\log_2(e|\widetilde{\mathscr{V}}_{\mathcal{H}}|)}{|\widetilde{\mathscr{V}}_{\mathcal{H}}|^{1-8Lc^4}}\\
&\le \eta^2+\frac{\eta^2}{2\log 2}\cdot\frac{\log(|\widetilde{\mathscr{V}}_{\mathcal{H}}|)}{|\widetilde{\mathscr{V}}_{\mathcal{H}}|^{1-8Lc^4}}\\
&\le \eta^2+\frac{\eta^2}{2\log 2}\\
&\le 2\eta^2.
\end{aligned}
$$

注意，我们使用了 $\mathscr{A}_{\mathcal{H}}\le L\log(e|\widetilde{\mathscr{V}}_{\mathcal{H}}|)$。我们还结合不等式 $\log(x)\le\sqrt{x}$，对所有 $x>0$ 都成立，得到了倒数第二行。因此，我们已经证明

$$
\chi^2(P_\pi\,||\,P_0)\le 2\eta^2+\left(2\eta^2+1\right)-1=4\eta^2.
$$

代入（37）可得

$$
\inf_{\varphi}\left\{P_0\{\varphi=1\}+\max_{s\ge\sqrt{p\mathscr{A}_{\mathcal{H}}}}\sup_{\substack{\Theta\in\mathscr{T}_s,\\ ||\Theta||_F\ge c\psi_{\text{adapt}}(p,s,n)}}P_\Theta\{\varphi=0\}\right\}\ge 1-\eta.
$$

证明完毕。 □

7.4 自适应上界

Proof of Theorem 3. 固定 $\eta\in(0,1)$。在证明的各处，我们会指出，可以将 C_η 取得足够大，以获得所需的界，因此暂时令 $C>C_\eta$。令 L^* 表示引理 21中的通用常数。令 D 表示以下各项中相应通用常数的最大值

引理 15 和11。令

$$K_2 := 1 \vee \frac{1}{(\log 2)^{1/4}} \vee \left(\frac{2}{c}\right)^{1/4} \vee \left(\frac{L^*}{c}\right)^{1/4},$$
$$K_3 := \frac{L^*}{K_2^2},$$
$$K_2' := \frac{1}{\sqrt{\log 2}} \vee \left(\frac{2}{c'}\right)^{1/2} \vee (c' K_3^2)^{-1/4}.$$

这里， $c := c^* \wedge c^{**}$ ，其中 c^* 和 c^{**} 分别是引理15和引理14的指数项中的通用常数。同理， $c' := (c')^* \wedge (c')^{**}$ ，其中 $(c')^*$ 和 $(c')^{**}$ 分别是引理11和引理10的指数项中的通用常数。请注意，这里对 K_2、K_3、K_2' 的取值与命题2和命题3证明中的取值几乎完全相同。唯一的修改是项 $(2/c)^{1/4}$ 和 $(2/c')^{1/4}$；这一修改的作用将在证明过程中逐渐显现。

对于任意 ν，定义

$$\mathcal{S}_{\text{bulk}} := \left\{ s \in \mathscr{S} : s < \sqrt{p\mathscr{A}_{\mathcal{H}}} \text{ and } \sqrt{\log\left(1 + \frac{p\mathscr{A}_{\mathcal{H}}}{s^2}\right)} \le K_3 \sqrt{d_\nu} \right\},$$
$$\mathcal{S}_{\text{tail}} := \left\{ s \in \mathscr{S} : s < \sqrt{p\mathscr{A}_{\mathcal{H}}} \text{ and } \sqrt{\log\left(1 + \frac{p\mathscr{A}_{\mathcal{H}}}{s^2}\right)} > K_3 \sqrt{d_\nu} \right\}.$$

我们分别考察第一类错误和第二类错误。先看第一类错误，由并集界可得

$$P_0\left\{\max_{\nu \in \mathscr{V}_{\mathcal{H}}} \max_{s \in \mathscr{S}} \varphi_{\nu,s} = 1\right\}$$
$$\le \left(\sum_{\nu \in \mathscr{V}_{\mathcal{H}}} \sum_{s \in \mathcal{S}_{\text{bulk}}} P_0\{\varphi_{\nu,s} = 1\}\right) + \left(\sum_{\nu \in \mathscr{V}_{\mathcal{H}}} \sum_{s \in \mathcal{S}_{\text{tail}}} P_0\{\varphi_{\nu,s} = 1\}\right) + \left(\sum_{\nu \in \mathscr{V}_{\mathcal{H}}} P_0\{\varphi_{\nu,p} = 1\}\right). \quad (39)$$

我们分别对每一项进行界定。

第一类错误：主体部分

对于 $s \in \mathcal{S}_{\text{bulk}}$,

$$P_0\{\varphi_{\nu,s} = 1\} = P_0\left\{T_{r_{\nu,s}}(d_\nu) > Cs\sqrt{\nu \log\left(1 + \frac{p\mathscr{A}_{\mathcal{H}}}{s^2}\right)}\right\}.$$

注意，引理 15 给出了以下结论。如果我们选择 $x > 0$ 满足

$$C^*\left(\sqrt{p r_{\nu,s}^4 e^{-c^* \frac{r_{\nu,s}^2}{d}} x} + \frac{d_\nu}{r_{\nu,s}^2} x\right) \le Cs\sqrt{\nu \log\left(1 + \frac{p\mathscr{A}_{\mathcal{H}}}{s^2}\right)}, \quad (40)$$

于是我们有 $P_0\{\varphi_{\nu,s} = 1\} \le e^{-x}$。接下来选择

$$x = \frac{C}{4(C^* \vee (C^*)^2)} \cdot \frac{1}{(2K_2^4 \lceil D \rceil) \vee (\sqrt{2\lceil D \rceil}/K_2^2)} \left(\frac{p\mathscr{A}_{\mathcal{H}}^2}{s^2} \wedge s \log\left(1 + \frac{p\mathscr{A}_{\mathcal{H}}}{s^2}\right)\right).$$

为验证（40），注意到 $c^* K_2^4 \geq 2$ 根据我们对 K_2 的选择，因此

$$\begin{aligned}\sqrt{pr_{\nu,s}^4 e^{-c^* \frac{r_{\nu,s}^2}{d}} x} &= \sqrt{pK_2^4 d_\nu \log\left(1+\frac{p\mathscr{A}_\mathcal{H}}{s^2}\right) \frac{s^4}{(s^2+p\mathscr{A}_\mathcal{H})^2} x} \\ &\leq \frac{\sqrt{C}}{2C^*} \frac{1}{\sqrt{2\lceil D\rceil}} s\sqrt{d_\nu \log\left(1+\frac{p\mathscr{A}_\mathcal{H}}{s^2}\right)} \\ &\leq \frac{\sqrt{C}}{2C^*} s\sqrt{\nu \log\left(1+\frac{p\mathscr{A}_\mathcal{H}}{s^2}\right)}.\end{aligned}$$

这里我们用到了 $d_\nu = \nu \vee \lceil D\rceil$。我们还用到了 $C \geq 1$；只要选择的 C_η 足够大（即 $C_\eta \geq 1$），该结论就成立。

同样，考虑

$$\begin{aligned}\frac{d_\nu}{r_{\nu,s}^2} x &= \frac{d_\nu}{K_2^2 \sqrt{d_\nu \log\left(1+\frac{p\mathscr{A}_\mathcal{H}}{s^2}\right)}} x \\ &\leq \frac{C}{4\sqrt{2\lceil D\rceil} C^*} s\sqrt{d_\nu \log\left(1+\frac{p\mathscr{A}_\mathcal{H}}{s^2}\right)} \\ &\leq \frac{C}{4C^*} s\sqrt{\nu \log\left(1+\frac{p\mathscr{A}_\mathcal{H}}{s^2}\right)}.\end{aligned}$$

因此，(40) 成立，于是我们有

$$P_0\left\{\varphi_{\nu,s}=1\right\} \leq \exp\left(-C\kappa\left(\frac{p\mathscr{A}_\mathcal{H}^2}{s^2} \wedge s\log\left(1+\frac{p\mathscr{A}_\mathcal{H}}{s^2}\right)\right)\right)$$

其中 κ 是通用常数 $\kappa = \frac{1}{4\ C^* \vee ((C^*)^2)}$。~~$\frac{1}{(2K_2^4\lceil D\rceil)\vee(\sqrt{2}/K_2^2)}$。有了~~这个界，注意到

$$\begin{aligned}&\sum_{\nu\in\mathscr{V}_\mathcal{H}} \sum_{s\in\mathcal{S}_{\text{bulk}}} P_0\left\{\varphi_{\nu,s}=1\right\} \\ &\leq \sum_{\nu\in\mathscr{V}_\mathcal{H}} \sum_{s\in\mathcal{S}_{\text{bulk}}} \exp\left(-C\kappa\left(\frac{p\mathscr{A}_\mathcal{H}^2}{s^2} \wedge s\log\left(1+\frac{p\mathscr{A}_\mathcal{H}}{s^2}\right)\right)\right) \\ &\leq |\mathscr{V}_\mathcal{H}| \sum_{\substack{k\in\mathbb{N}\cup\{0\}:\\ 2^k<\sqrt{p\mathscr{A}_\mathcal{H}}}} \exp\left(-C\kappa\left(\frac{p\mathscr{A}_\mathcal{H}^2}{2^{2k}} \wedge 2^k\log\left(1+\frac{p\mathscr{A}_\mathcal{H}}{2^{2k}}\right)\right)\right) \\ &\leq e^{2\mathscr{A}_\mathcal{H}} \sum_{\substack{k\in\mathbb{N}\cup\{0\}:\\ 2^k<\sqrt{p\mathscr{A}_\mathcal{H}}}} \exp\left(-C\kappa\frac{p\mathscr{A}_\mathcal{H}^2}{2^{2k}}\right) + e^{2\mathscr{A}_\mathcal{H}} \sum_{\substack{k\in\mathbb{N}\cup\{0\}:\\ 2^k<\sqrt{p\mathscr{A}_\mathcal{H}}}} \exp\left(-C\kappa 2^k\log\left(1+\frac{p\mathscr{A}_\mathcal{H}}{2^{2k}}\right)\right).\end{aligned}$$

这里我们用到了引理 3中的 $\log(e|\mathscr{V}_\mathcal{H}|) \leq 2\mathscr{A}_\mathcal{H}$。我们分别对每一项进行估计。首先，考虑

$$\begin{aligned}e^{2\mathscr{A}_\mathcal{H}} \sum_{\substack{k\in\mathbb{N}\cup\{0\}:\\ 2^k<\sqrt{p\mathscr{A}_\mathcal{H}}}} \exp\left(-C\kappa\frac{p\mathscr{A}_\mathcal{H}^2}{2^{2k}}\right) &= e^{2\mathscr{A}_\mathcal{H}} \sum_{\substack{k\in\mathbb{N}:\\ 2^k<\sqrt{p\mathscr{A}_\mathcal{H}}}} \exp\left(-C\kappa\mathscr{A}_\mathcal{H}\cdot\left(\frac{\sqrt{p\mathscr{A}_\mathcal{H}}}{2^k}\right)^2\right) \\ &\leq \sum_{k=0}^{\infty} \exp\left(-(C\kappa-2)\mathscr{A}_\mathcal{H} 4^k\right) \\ &\leq \frac{\eta}{12}\end{aligned}$$

只要我们取 C_η 足够大。

同样，考虑

$$e^{2\mathscr{A}_{\mathcal{H}}} \sum_{\substack{k\in\mathbb{N}\cup\{0\}:\\ 2^k<\sqrt{p\mathscr{A}_{\mathcal{H}}}}} \exp\left(-C\kappa 2^k \log\left(1+\frac{p\mathscr{A}_{\mathcal{H}}}{2^{2k}}\right)\right) \le e^{2\mathscr{A}_{\mathcal{H}}} \sum_{\substack{k\in\mathbb{N}\cup\{0\}:\\ 2^k<\sqrt{p\mathscr{A}_{\mathcal{H}}}}} \exp\left(-\frac{C\kappa}{2} 2^k \log\left(1+\frac{p\mathscr{A}_{\mathcal{H}}}{2^{2k}}\right) - \frac{C\kappa}{2}\mathscr{A}_{\mathcal{H}}\right)$$

$$\le e^{-\left(\frac{C\kappa}{2}-2\right)\mathscr{A}_{\mathcal{H}}} \sum_{\substack{k\in\mathbb{N}\cup\{0\}:\\ 2^k<\sqrt{p\mathscr{A}_{\mathcal{H}}}}} \exp\left(-\frac{C\kappa}{2} 2^k \log\left(1+\frac{p\mathscr{A}_{\mathcal{H}}}{2^{2k}}\right)\right)$$

$$\le \frac{\eta}{12}$$

同样，只要我们取 C_η 足够大。我们利用了 $\mathscr{A}_{\mathcal{H}}$ 关于 p至多呈对数增长这一点，即使用粗略界 $\mathscr{A}_{\mathcal{H}} \le \log(ep)$。因此，我们得出

$$\sum_{\nu\in\mathcal{V}_{\mathcal{H}}} \sum_{s\in\mathcal{S}_{\text{bulk}}} P_0\left\{\varphi_{\nu,s}=1\right\} \le \frac{\eta}{6}. \tag{41}$$

第一类错误：尾部

对于 $s\in\mathcal{S}_{\text{tail}}$，我们有

$$P_0\left\{\varphi_{\nu,s}=1\right\} = P_0\left\{T_{r'_s}(d_\nu) > Cs\log\left(1+\frac{p\mathscr{A}_{\mathcal{H}}}{s^2}\right)\right\}.$$

注意，引理 11 给出了与（40）类似的结论，即如果我们选择 $x>0$ 满足

$$C^{**}\left(\sqrt{p(r'_s)^4 e^{-c^{**}(r'_s)^2}x} + x\right) \le Cn\tau^2_{\text{adapt}}(p,s,n)^2, \tag{42}$$

那么我们有 $P_0\left\{\varphi_{\nu,s}=1\right\} \le e^{-x}$。令我们选择

$$x = \frac{C}{4}\cdot\frac{1}{2\left((C^{**})\vee(C^{**})^2\right)\left((K'_2)^4\vee 1\right)}\left(\frac{p\mathscr{A}_{\mathcal{H}}^2}{s^2}\wedge s\log\left(1+\frac{p\mathscr{A}_{\mathcal{H}}}{s^2}\right)\right).$$

为验证（42），注意到 $c^{**}\left(K'_2\right)^2 \ge 2$ 根据我们对 K'_2的选择，因此

$$\sqrt{p(r'_s)^4 e^{-c^{**}(r'_s)^2}x} = (K'_2)^2\log\left(1+\frac{p\mathscr{A}_{\mathcal{H}}}{s^2}\right)\sqrt{p\frac{s^4}{(s^2+p\mathscr{A}_{\mathcal{H}})^2}x}$$

$$\le \sqrt{C}\frac{1}{2C^{**}}\log\left(1+\frac{p\mathscr{A}_{\mathcal{H}}}{s^2}\right)\sqrt{p\frac{s^4}{(s^2+p\mathscr{A}_{\mathcal{H}})^2}\cdot\frac{p\mathscr{A}_{\mathcal{H}}^2}{s^2}}$$

$$\le \frac{C}{2C^{**}}s\log\left(1+\frac{p\mathscr{A}_{\mathcal{H}}}{s^2}\right).$$

这里，我们使用了 $C\ge 1$。同样，考虑

$$x \le \frac{C}{2C^{**}}\cdot s\log\left(1+\frac{p\mathscr{A}_{\mathcal{H}}}{s^2}\right).$$

因此，(42) 得到满足，从而我们有

$$P_0\left\{\varphi_{\nu,s}=1\right\} \le \exp\left(-C\kappa\left(\frac{p\mathscr{A}_{\mathcal{H}}^2}{s^2}\wedge s\log\left(1+\frac{p\mathscr{A}_{\mathcal{H}}}{s^2}\right)\right)\right)$$

其中 κ 是通用常数 $\kappa = \frac{1}{4} \cdot {}_{2}^{1}$ （~~$(C^{**})(C^{**})^2$ 。有~~了这个界限，我们注意到

$$
\begin{aligned}
&\sum_{\nu \in \mathscr{V}_{\mathcal{H}}} \sum_{s \in \mathcal{S}_{\text{tail}}} P_0\left\{\varphi_{\nu,s} = 1\right\} \\
&\leq \sum_{\nu \in \mathscr{V}_{\mathcal{H}}} \sum_{s \in \mathcal{S}_{\text{tail}}} \exp\left(-C\kappa\left(\frac{p\mathscr{A}_{\mathcal{H}}^2}{s^2} \wedge s\log\left(1 + \frac{p\mathscr{A}_{\mathcal{H}}}{s^2}\right)\right)\right) \\
&\leq |\mathscr{V}_{\mathcal{H}}| \sum_{\substack{k \in \mathbb{N} \cup \{0\}: \\ 2^k < \sqrt{p\mathscr{A}_{\mathcal{H}}}}} \exp\left(-C\kappa\left(\frac{p\mathscr{A}_{\mathcal{H}}^2}{2^{2k}} \wedge 2^k \log\left(1 + \frac{p\mathscr{A}_{\mathcal{H}}}{2^{2k}}\right)\right)\right) \\
&\leq e^{2\mathscr{A}_{\mathcal{H}}} \sum_{\substack{k \in \mathbb{N} \cup \{0\}: \\ 2^k < \sqrt{p\mathscr{A}_{\mathcal{H}}}}} \exp\left(-C\kappa\frac{p\mathscr{A}_{\mathcal{H}}^2}{2^{2k}}\right) + e^{2\mathscr{A}_{\mathcal{H}}} \sum_{\substack{k \in \mathbb{N} \cup \{0\}: \\ 2^k < \sqrt{p\mathscr{A}_{\mathcal{H}}}}} \exp\left(-C\kappa 2^k \log\left(1 + \frac{p\mathscr{A}_{\mathcal{H}}}{2^{2k}}\right)\right).
\end{aligned}
$$

这里我们使用了引理 $e|\mathscr{V}_{\mathcal{H}}|) \leq 2\mathscr{A}_{\mathcal{H}}$ 中的 log(3。由此，可以采用稠密情形中的相同论证，得出

$$
\sum_{\nu \in \mathscr{V}_{\mathcal{H}}} \sum_{s \in \mathcal{S}_{\text{tail}}} P_0\left\{\varphi_{\nu,s} = 1\right\} \leq \frac{\eta}{6} \tag{43}
$$

只要 C_η 足够大即可。

第一类错误：稠密

现在我们来界定 $\sum_{\nu \in \mathscr{V}_{\mathcal{H}}} P_0\left\{\varphi_{\nu,p} = 1\right\}$。注意，对任意 $\nu \in \mathscr{V}_{\mathcal{H}}$，都有

$$
\begin{aligned}
P_0\left\{\varphi_{\nu,p} = 1\right\} &= P_0\left\{n\sum_{j=1}^{p}\sum_{k \leq \nu} X_{k,j}^2 > C\sqrt{\nu p\mathscr{A}_{\mathcal{H}}}\right\} \\
&= P\left\{\chi_{\nu p}^2 > C\sqrt{\nu p\mathscr{A}_{\mathcal{H}}}\right\} \\
&\leq P\left\{\chi_{\nu p}^2 > \frac{C}{2}\left(\sqrt{\nu p\mathscr{A}_{\mathcal{H}}} + \mathscr{A}_{\mathcal{H}}\right)\right\} \\
&\leq P\left\{\chi_{\nu p}^2 > 2\sqrt{\nu p \cdot \frac{C}{4}\mathscr{A}_{\mathcal{H}}} + 2 \cdot \frac{C}{4}\mathscr{A}_{\mathcal{H}}\right\} \\
&\leq e^{-\frac{C}{4}\mathscr{A}_{\mathcal{H}}}.
\end{aligned}
$$

为得到第四行，我们使用了 $\mathscr{A}_{\mathcal{H}} \leq \log(ep)$，这意味着 $\sqrt{\nu p\mathscr{A}_{\mathcal{H}}} \geq \mathscr{A}_{\mathcal{H}}$。在上述推导中，我们还使用了 $\frac{C}{2} \geq 1$ （只要取 C_η 足够大，该结论便成立）来得到倒数第二行，并使用引理 16 得到最后一行。因此，

$$
\sum_{\nu \in \mathscr{V}_{\mathcal{H}}} P_0\left\{\varphi_{\nu,p} = 1\right\} \leq |\mathscr{V}_{\mathcal{H}}| e^{-\frac{C}{4}\mathscr{A}_{\mathcal{H}}} \leq e^{-(\frac{C}{4}-2)\mathscr{A}_{\mathcal{H}}} \leq \frac{\eta}{6} \tag{44}
$$

由于根据引理 $e|\mathscr{V}_{\mathcal{H}}|) \leq 2\mathscr{A}_{\mathcal{H}}$ 有 log(3 ，并且 $\frac{C}{4} - 2 \geq \log(6/\eta)$，只要取 C_η足够大，该结论便成立。

将 (41)、(43)、(44) 合并到 (39) 中，可得第一类错误界

$$
P_0\left\{\max_{\nu \in \mathscr{V}_{\mathcal{H}}} \max_{s \in \mathscr{S}} \varphi_{\nu,s} = 1\right\} \leq \frac{\eta}{6} + \frac{\eta}{6} + \frac{\eta}{6} = \frac{\eta}{2}. \tag{45}
$$

第二类错误：现在我们考察第二类错误。假设 $s^* \in [p]$。令 $f \in \mathcal{F}_{s^*}$ with$||f||_2 \geq C\tau_{\text{adapt}}(p, s^*, n)$。接下来我们考虑各种情形。

第二类错误：稠密

假设 $s^* \geq \sqrt{p\mathscr{A}_{\mathcal{H}}}$。令 $\tilde{\nu}$ 表示 $\mathscr{V}_{\mathcal{H}}$ 中大于或等于

$\nu_{\mathcal{H}}(s^*, \mathscr{A}_{\mathcal{H}})$。则

$$P_f\left\{\max_{\nu\in\mathscr{V}_{\mathcal{H}}}\max_{s\in\mathscr{S}}\varphi_{\nu,s}=0\right\}\leq P_f\left\{\varphi_{\tilde{\nu},p}=0\right\}=P_f\left\{n\sum_{j=1}^{p}\sum_{k\leq\tilde{\nu}}X_{k,j}^2>\tilde{\nu}p+C\sqrt{\tilde{\nu}p\mathscr{A}_{\mathcal{H}}}\right\}.$$

考虑

$$n\sum_{j=1}^{p}\sum_{k\leq\tilde{\nu}}X_{k,j}^2\sim\chi^2_{p\tilde{\nu}}\left(n\sum_{j=1}^{p}\sum_{k\leq\tilde{\nu}}\theta_{k,j}^2\right)$$

其中，集合 $\{\theta_{k,j}\}$ 表示 f 的基系数集合。我们还将使用矩阵 Θ 表示基系数集合；注意， $\Theta\in\mathscr{T}_{s^*}$。令 S^* 表示 j 的集合，其中 Θ_j 非零。注意

$$\begin{aligned}
C^2\tau^2_{\text{adapt}}(p,s^*,n)&\leq||f||_2^2\\
&=\sum_{j=1}^{p}\sum_{k=1}^{\infty}\theta_{k,j}^2\\
&\leq\sum_{j\in S^*}\left(\sum_{k\leq\tilde{\nu}}\theta_{k,j}^2+\sum_{k>\tilde{\nu}}\theta_{k,j}^2\right)\\
&\leq\sum_{j\in S^*}\left(\sum_{k\leq\tilde{\nu}}\theta_{k,j}^2+\mu_{\tilde{\nu}}\sum_{k>\tilde{\nu}}\frac{\theta_{k,j}^2}{\mu_k}\right)\\
&\leq\left(\sum_{j\in S^*}\sum_{k\leq\tilde{\nu}}\theta_{k,j}^2\right)+s^*\mu_{\tilde{\nu}}\\
&=\left(\sum_{j=1}^{p}\sum_{k\leq\tilde{\nu}}\theta_{k,j}^2\right)+s^*\mu_{\tilde{\nu}}.
\end{aligned}$$

因此，

$$\begin{aligned}
n\sum_{j=1}^{p}\sum_{k\leq\tilde{\nu}}\theta_{k,j}^2&\geq C^2n\tau^2_{\text{adapt}}(p,s^*,n)-ns^*\mu_{\tilde{\nu}}\\
&\geq C^2n\tau^2_{\text{adapt}}(p,s^*,n)-s^*\sqrt{\tilde{\nu}\log\left(1+\frac{p\mathscr{A}_{\mathcal{H}}}{(s^*)^2}\right)}\\
&\geq C^2n\tau^2_{\text{adapt}}(p,s^*,n)-\sqrt{2}n\tau^2_{\text{adapt}}(p,s^*,n)\\
&\geq(C^2-\sqrt{2})n\tau^2_{\text{adapt}}(p,s^*,n)
\end{aligned}$$

我们利用了 $\tilde{\nu}\leq 2\nu_{\mathcal{H}}(s^*,\mathscr{A}_{\mathcal{H}})$，从而得到第三行。因此，我们有

$$n\sum_{j=1}^{p}\sum_{k\leq\tilde{\nu}}\theta_{k,j}^2\geq\left(C^2-\sqrt{2}\right)n\tau^2_{\text{adapt}}(p,s^*,n)\geq\frac{C^2-\sqrt{2}}{\sqrt{2}}\sqrt{\tilde{\nu}p\mathscr{A}_{\mathcal{H}}}.$$

信号幅度满足成功检测所需的强度条件，这一点可参照命题的论证得出 4。因此，通过选取足够大的 $\frac{\eta}{2}$，我们可以使第二类错误小于或等于 C_η。将此界与（39）结合，可得检验风险总量不超过 η，如所求。

第二类错误：主体部分
假设 $s^*<\sqrt{p\mathscr{A}_{\mathcal{H}}}$ 和 $\sqrt{\log\left(1+\frac{p\mathscr{A}_{\mathcal{H}}}{s^*}()^2\right)}\leq K_3\sqrt{\nu_{\mathcal{H}}(s^*,\mathscr{A}_{\mathcal{H}})\vee\lceil D\rceil_0}$ 令 $\tilde{\nu}$ 为 $\mathscr{V}_{\mathcal{H}}$ 中大于或等于 $\nu_{\mathcal{H}}(s^*,\mathscr{A}_{\mathcal{H}})$ 的最小元素。令 $\tilde{s}$ 为 $\mathscr{S}$ 中大于或等于 s^* 的最小元素。根据这些网格的定义，我们有 $\tilde{\nu}/2<\nu_{\mathcal{H}}(s^*,\mathscr{A}_{\mathcal{H}})\leq\tilde{\nu}$ 和 $s^*\leq\tilde{s}\leq 2s^*$。

Then

$$P_f\left\{\max_{\nu\in\mathscr{V}_{\mathcal{H}}}\max_{s\in\mathscr{S}}\varphi_{\nu,s}=0\right\}\le P_f\left\{\varphi_{\tilde{\nu},\tilde{s}}=0\right\}.$$

采用这些选择后，我们有

$$\sqrt{\log\left(1+\frac{p\mathscr{A}_{\mathcal{H}}}{\tilde{s}^2}\right)}\le K_3\sqrt{d_{\tilde{\nu}}}.$$

注意，由于 $\tilde{s}\ge s^*$，我们有 $f\in\mathcal{F}_{\tilde{s}}$。按照命题 2 证明中的类似论证可知，要成功检测所需的信号强度平方阶为

$$\frac{\tilde{s}\sqrt{\tilde{\nu}\log\left(1+\frac{p\mathscr{A}_{\mathcal{H}}}{\tilde{s}^2}\right)}}{n}\asymp\frac{s^*\sqrt{\nu_{\mathcal{H}}(s^*,\mathscr{A}_{\mathcal{H}})\log\left(1+\frac{p\mathscr{A}_{\mathcal{H}}}{(s^*)^2}\right)}}{n}.$$

这恰好是 $\tau^2_{\text{adapt}}(p,\ s^*,\ n)$ 的阶，因此通过选择 $\frac{\eta}{2}$ 足够大的 C_η，可以使第二类错误小于 $\frac{\eta}{2}$。将此界与（39）结合，可得总检验风险不超过 η，如所求。

第二类错误：尾部

假设 $s^*<\sqrt{p\mathscr{A}_{\mathcal{H}}}$ 和 $\sqrt{\log\left(1+\frac{p\mathscr{A}_{\mathcal{H}}}{s^*()^2}\right)}>K_3\sqrt{\nu_{\mathcal{H}}(s^*,\mathscr{A}_{\mathcal{H}})\vee\lceil D\rceil}$。令 $\tilde{\nu}$ 和 $\tilde{s}$ 如前一情形中所定义，即稠密情形下的第二类错误分析。与前面一样，我们有

$$P_f\left\{\max_{\nu\in\mathscr{V}_{\mathcal{H}}}\max_{s\in\mathscr{S}}\varphi_{\nu,s}=0\right\}\le P_f\left\{\varphi_{\tilde{\nu},\tilde{s}}=0\right\}.$$

现在，考虑到 $s^*\le\tilde{s}$，因此 $f\in\mathcal{F}_{\tilde{s}}$。

假设我们有

$$\sqrt{\log\left(1+\frac{p\mathscr{A}_{\mathcal{H}}}{\tilde{s}^2}\right)}>K_3\sqrt{d_{\tilde{\nu}}}.$$

我们可以采用与命题证明中类似的论证 3，得出要成功检测所需的信号强度为平方量级

$$\frac{\tilde{s}\log\left(1+\frac{p\mathscr{A}_{\mathcal{H}}}{\tilde{s}^2}\right)}{n}\asymp\frac{s^*\log\left(1+\frac{p\mathscr{A}_{\mathcal{H}}}{(s^*)^2}\right)}{n}.$$

这恰好是 $\tau^2_{\text{adapt}}(p,s^*,n)$ 的量级，因此通过选择足够大的 C_η，我们可以使第二类错误小于 $\frac{\eta}{2}$。

另一方面，假设我们有

$$\sqrt{\log\left(1+\frac{p\mathscr{A}_{\mathcal{H}}}{\tilde{s}^2}\right)}\le K_3\sqrt{d_{\tilde{\nu}}}$$

正如上一节关于主体区域的讨论所述，要成功检测所需的信号强度为平方量级

$$\frac{\tilde{s}\sqrt{\tilde{\nu}\log\left(1+\frac{p\mathscr{A}_{\mathcal{H}}}{\tilde{s}^2}\right)}}{n}\asymp\frac{s^*\sqrt{\nu_{\mathcal{H}}(s^*,\mathscr{A}_{\mathcal{H}})\log\left(1+\frac{p\mathscr{A}_{\mathcal{H}}}{(s^*)^2}\right)}}{n}.$$

考虑到

$$\tau^2_{\text{adapt}}(p,s^*,n)=\frac{s^*\log\left(1+\frac{p\mathscr{A}_{\mathcal{H}}}{(s^*)^2}\right)}{n}\gtrsim\frac{s^*\sqrt{\nu_{\mathcal{H}}(s^*,\mathscr{A}_{\mathcal{H}})\log\left(1+\frac{p\mathscr{A}_{\mathcal{H}}}{(s^*)^2}\right)}}{n}$$

因此，通过选择 $\frac{\eta}{2}$ 足够大，我们可以得到第二类错误小于 C_η。将此界与（39）结合，可得总检验风险的上界为 η，如所需。 □

7.5 光滑性与稀疏性的自适应下界

Proof of Theorem 5. 固定 $\eta \in (0,1)$。任取 $s \geq p^{1/2+\delta} \sqrt{\log \log n}$。在证明过程中，我们会说明何时必须令 c_η 足够小，因此暂令 $0 < c < c_\eta$。对于每个 s，定义几何网格

$$\mathcal{V}_s := \left\{ 2^k : k \in \mathbb{N} \text{ and } \left(\frac{ns}{\sqrt{p \log \log(np)}} \right)^{\frac{2}{4\alpha_1+1}} \leq 2^k \leq \left(\frac{ns}{\sqrt{p \log \log(np)}} \right)^{\frac{2}{4\alpha_0+1}} \right\}.$$

注意，$\log |\mathcal{V}_s| \asymp \log \log(np)$。

现在我们定义一个先验 π，其支撑集位于备择假设上。令 $\nu \sim \text{Uniform}(\mathcal{V}_s)$. 令 $\alpha = \alpha(\nu, s)$ 表示下式的解

$$\nu = \left(\frac{ns}{\sqrt{p \log \log(np)}} \right)^{\frac{2}{4\alpha+1}}.$$

注意，对于任意 $\nu \in \mathcal{V}_s$，$\alpha(\nu, s) \in [\alpha_0, \alpha_1]$。注意，$\alpha$ 是随机的，因为 ν 是随机的。从大小为 s 的集合中均匀随机抽取一个子集 $S \subset [p]$。独立地抽取

$$\theta_{k,j} \sim \begin{cases} \text{Uniform}\{-c\rho_\nu, c\rho_\nu\} & \text{if } j \in S \text{ and } k \leq \nu, \\ \delta_0 & \text{otherwise.} \end{cases}$$

这里，ρ_ν 由 $\rho_\nu := \left(\sqrt{p} \right)^{-\frac{2\alpha+1}{4\alpha+1}} \left(\frac{ns}{\log \log(np)} \right)$ 给出。定义了 ρ_ν 之后，先验 π 的定义便完整了。

现在我们必须证明 π 确实支持备择假设。注意，对于 $\Theta \sim \pi$，我们有

$$||\Theta||_F^2 = c^2 s \rho_\nu^2 \nu = c^2 s \left(\frac{ns}{\sqrt{p \log \log(np)}} \right)^{-\frac{4\alpha}{4\alpha+1}} = c^2 \tau_{\text{dense}}^2(p, s, n, \alpha).$$

此外，考虑到对于 $i \in S$，根据 $\alpha = \alpha(\nu, s)$ 的定义，

$$\sum_{\ell=1}^{\infty} \frac{\theta_{i,\ell}^2}{\ell^{-2\alpha}} = c^2 \rho_\nu^2 \nu^{2\alpha} \sum_{\ell \leq \nu} \frac{\ell^{2\alpha}}{\nu^{2\alpha}} \leq c^2 \rho_\nu^2 \nu^{2\alpha+1} = c^2 \rho_\nu^2 \left(\frac{ns}{\sqrt{p \log \log(np)}} \right)^{\frac{4\alpha+2}{4\alpha+1}} \leq c^2 \leq 1.$$

我们可以取 $c_\eta \leq 1$ 以确保 $c^2 \leq 1$。当然，对于 $i \in S^c$，我们有 $\Theta_i = 0$。因此，我们已经证明$\Theta \in \mathcal{F}(s, \alpha)$以概率1成立，所以 π 具有正确的支撑。

记 $P_\pi = \int P_\Theta \pi(d\Theta)$ 为由先验 π 诱导的混合分布，我们有

$$\inf_\varphi \left\{ P_0\{\varphi \neq 0\} + \max_{\tilde{s} \geq p^{1/2+\delta}} \sup_{\alpha \in [\alpha_0, \alpha_1]} \sup_{\substack{f \in \mathcal{F}(\tilde{s}, \alpha), \\ ||f||_2 \geq c\tau_{\text{dense}}(p, \tilde{s}, n, \alpha)}} P_f\{\varphi \neq 1\} \right\} \geq 1 - \frac{1}{2} \sqrt{\chi^2(P_\pi \,||\, P_0)}.$$

在以下计算中，令 Θ、$\Theta' \overset{iid}{\sim} \pi$。令 ν、$\nu' \in \mathcal{V}_s$ 为相应的量，并令 α、α' 表示相应的光滑度级别。进一步令 S、S' 表示相应的支撑集。注意，二者的大小均为 s。令 $\{r_{i\ell}, r'_{i\ell}\}_{1 \leq i \leq p,\ \ell \in \mathbb{N}}$ 表示一个由 Rademacher(1/2) 随机变量组成的独立同分布集合，该集合与所有其他随机变量相互独立。根据 Ingster-Suslina

方法（命题 5），我们有

$$\begin{aligned}\chi^2(P_\pi \,||\, P_0)+1 &= E\left(\exp\left(n\langle\Theta,\Theta'\rangle_F\right)\right)\\ &= E\left(\exp\left(c^2 n\rho_\nu\rho_{\nu'}\sum_{i\in S\cap S'}\sum_{\ell=1}^{\nu\wedge\nu'} r_{i\ell}r'_{i\ell}\right)\right)\\ &\le E\left(\exp\left(\frac{c^4}{2}n^2\rho_\nu^2\rho_{\nu'}^2(\nu\wedge\nu')|S\cap S'|\right)\right)\\ &\le E\left(\exp\left(\frac{c^4}{2}n^2\sqrt{\rho_\nu^4\nu\cdot\rho_{\nu'}^4\nu'}\cdot\frac{\nu\wedge\nu'}{\sqrt{\nu\nu'}}|S\cap S'|\right)\right).\end{aligned}$$

根据 $\alpha(\nu,s)$ 的定义，我们有

$$\begin{aligned}\sqrt{\rho_\nu^4\nu\cdot\rho_{\nu'}^4\nu'} &= \left(\left(\frac{ns}{\sqrt{p\log\log(np)}}\right)^{-\frac{8\alpha+4}{4\alpha+1}+\frac{2}{4\alpha+1}}\left(\frac{ns}{\sqrt{p\log\log(np)}}\right)^{-\frac{8\alpha'+4}{4\alpha'+1}+\frac{2}{4\alpha'+1}}\right)^{1/2}\\ &= \left(\frac{\sqrt{p\log\log(np)}}{ns}\right)^2\\ &\le \frac{2\log\left(1+\frac{p\log\log(np)}{s^2}\right)}{n^2}.\end{aligned}$$

这里，我们使用了 $s\ge\sqrt{p\log\log(np)}$，因为 $s\ge p^{1/2+\delta}$ 且 $s\ge\sqrt{p\log\log n}$。我们还对 $0\le x\le 1$ 使用了不等式 $x/2\le\log(1+x)$。由此可得

$$\begin{aligned}\chi^2(P_\pi \,||\, P_0)+1 &\le E\left(\exp\left(c^4\cdot\log\left(1+\frac{p\log\log(np)}{s^2}\right)\cdot\frac{\nu\wedge\nu'}{\sqrt{\nu\nu'}}|S\cap S'|\right)\right)\\ &\le E\left(\left(1-\frac{s}{p}+\frac{s}{p}e^{c^4\left(\frac{\nu\wedge\nu'}{\sqrt{\nu\nu'}}\right)\log\left(1+\frac{p\log\log(np)}{s^2}\right)}\right)^s\right)\\ &\le E\left(\exp\left(c^4\frac{\nu\wedge\nu'}{\sqrt{\nu\nu'}}\log\log(np)\right)\right).\end{aligned}$$

注意到我们可以写成 $\nu=2^k$，$\nu'=2^{k'}$ 并且 $|\mathcal{V}_s|\asymp\log(np)$，我们可以按照 [16] 中命题 4.2 的证明采用相同步骤。取 c_η 充分小，可得 $\chi^2(P_\pi \,||\, P_0)\le 4\eta^2$ 从而得到

$$\inf_{\varphi}\left\{P_0\{\varphi\neq 0\}+\max_{\tilde{s}\ge p^{1/2+\delta}}\sup_{\alpha\in[\alpha_0,\alpha_1]}\sup_{\substack{f\in\mathcal{F}(\tilde{s},\alpha),\\ ||f||_2\ge c\tau_{\text{adapt}}(p,\tilde{s},n,\alpha)}}P_f\{\varphi\neq 1\}\right\}\ge 1-\eta$$

这正是所需。 □

Proof of Theorem 6. 为便于处理，我们作一些化简。注意，τ_{sparse} 可改写为

$$\tau^2_{\text{sparse}}(p,s,n,\alpha)\asymp\begin{cases}s\left(\frac{n}{\sqrt{\log(p\log\log n)}}\right)^{-\frac{4\alpha}{4\alpha+1}} & \text{if } \log(p\log\log n)\le n^{\frac{1}{2\alpha+1}},\\ \frac{s\log(p\log\log n)}{n} & \text{if } \log(p\log\log n)> n^{\frac{1}{2\alpha+1}}.\end{cases}$$

注意，在 $\log(p\log\log n)>n^{\frac{1}{2\alpha+1}}$ 的情形下，我们有 $\log p\gtrsim n^{\frac{1}{2\alpha+1}}$。因此，$\log(p\log\log n)\asymp\log p$，所以我们可以进一步化简为

$$\tau^2_{\text{sparse}}(p,s,n,\alpha)\asymp\begin{cases}s\left(\frac{n}{\sqrt{\log(p\log\log n)}}\right)^{-\frac{4\alpha}{4\alpha+1}} & \text{if } \log(p\log\log n)\le n^{\frac{1}{2\alpha+1}},\\ \frac{s\log p}{n} & \text{if } \log(p\log\log n)> n^{\frac{1}{2\alpha+1}}.\end{cases}\tag{46}$$

将 τ_{sparse} 写成 (46) 的形式很方便。下界 $s\log n$ 恰好就是极小极大下界，因此无需新的论证。只需证明下界 $s\left(n/\sqrt{\log(p\log\log n)}\right)^{-\frac{4\alpha}{4\alpha+1}}$。

证明与定理 5 的证明非常相似，因此为简洁起见，我们只指出其中的修改。任取 $s < p^{1/2-\delta}$。令 π 为定理 5 证明中的先验，但改用 $\mathcal{V}_{\text{sparse}}$ 由 (47) 给出，以及 $\alpha(\nu)$ 由 (48) 给出，具体定义如下。定义几何网格

$$\mathcal{V}_{\text{sparse}} := \left\{ 2^k : k \in \mathbb{N} \text{ and } \left(\frac{n}{\sqrt{\log(p \log\log n)}} \right)^{\frac{2}{4\alpha_1+1}} \le 2^k \le \left(\frac{n}{\sqrt{\log(p \log\log n)}} \right)^{\frac{2}{4\alpha_0+1}} \right\} \tag{47}$$

令 $\alpha(\nu)$ 表示方程的解

$$\nu = \left(\frac{n}{\sqrt{\log(p \log\log n)}} \right)^{\frac{2}{4\alpha+1}}. \tag{48}$$

注意，对于任意 $\nu \in \mathcal{V}_{\text{sparse}}$，$\alpha(\nu) \in [\alpha_0, \alpha_1]$。此外，使用

$$\rho_\nu = \left(\frac{n}{\sqrt{\log(p \log\log n)}} \right)^{-\frac{2\alpha+1}{4\alpha+1}}.$$

可以用同样的方法验证，π 经过这些修改后，在备择假设下得到了适当的支持。我们可以按照定理 5 的证明继续进行，直到得到

$$\sqrt{\rho_\nu^4 \nu \cdot \rho_{\nu'}^4 \nu'} = \left(\frac{\sqrt{\log(p \log\log n)}}{n} \right)^2.$$

从这里开始，我们利用这样一个事实：$s \le p^{\frac{1}{2}-\delta}$ 意味着 $\log(p \log\log n) \asymp \log\left(1 + \frac{p \log\log n}{s^2}\right)$。换句话说，存在一个常数 $\kappa > 0$，使得

$$\sqrt{\rho_\nu^4 \nu \cdot \rho_{\nu'}^4 \nu'} \le \kappa \frac{\log\left(1 + \frac{p \log\log n}{s^2}\right)}{n^2}.$$

然后取 c_η 充分小，便可继续定理 5 的其余证明，从而得到所需结果。

□

7.6 光滑性和稀疏性自适应上界

Proof of Theorem 4. 固定 $\eta \in (0, 1)$。为方便起见，我们直接记为 $\mathcal{V} = \mathcal{V}_{\text{test}}$。在证明过程中，我们会注明何时需要取 C_η 足够大，因此暂且令 $C > C_\eta$。我们也会注明何时需要取 K_η 足够大。我们还会注明何时需要取 K_η 足够大。我们先考察第一类错误。由并集界，并取 $K_\eta \ge 1$，

$$\begin{aligned}
P_0\left\{ \max_{\nu \in \mathcal{V}} \varphi_\nu = 1 \right\} &\le \sum_{\nu \in \mathcal{V}} P_0\{\varphi_\nu = 1\} \\
&= \sum_{\nu \in \mathcal{V}} P\left\{ \chi^2_{\nu p} \ge \nu p + K_\eta \left(\sqrt{\nu p \log\log(np)} + \log\log(np) \right) \right\} \\
&\le \sum_{\nu \in \mathcal{V}} P\left\{ \chi^2_{\nu p} \ge \nu p + 2\left(\sqrt{\nu p \frac{K_\eta}{4} \log\log(np)} + \frac{K_\eta}{4} \log\log(np) \right) \right\} \\
&\le |\mathcal{V}| e^{-\frac{K_\eta}{4} \log\log(np)} \\
&\le e^{-\left(\frac{K_\eta}{4} - \kappa\right) \log\log(np)}
\end{aligned}$$

对于某个普适正常数 κ。这里，我们使用了 $|\mathcal{V}| \asymp \log(np)$。适当增大 K_η，即可将第一类错误控制在适当范围内

$$P_0\left\{ \max_{\nu \in \mathcal{V}} \varphi_\nu = 1 \right\} \le \frac{\eta}{2}.$$

现在我们考察第二类错误。任取 $s^* \geq p^{1/2+\delta}$、 $\alpha^* \in [\alpha_0, \alpha_1]$和 $f^* \in \mathcal{F}(s^*, \alpha^*)$，其中 $||f^*||_2 \geq C\tau_{\text{dense}}(p, s^*, n, \alpha^*)$。令

$$\nu^* = \left(\frac{ns^*}{\sqrt{p \log\log(np)}} \right)^{\frac{2}{4\alpha^*+1}}.$$

令 $\nu \in \mathcal{V}$ 为大于 ν^*的最小元素。注意，根据 $\mathcal{V}$ 的定义， $\nu/2 \leq \nu^* \leq \nu$。

$$\begin{aligned}
&P_{f^*}\left\{\max_{\tilde{\nu}\in\mathcal{V}} \varphi_{\tilde{\nu}} = 0\right\} \\
&\leq P_{f^*}\{\varphi_\nu = 0\} \\
&= P_{f^*}\left\{\chi^2_{\nu p}(n||f^*||_2^2) \leq 2\sqrt{\nu p} + K_\eta\left(\sqrt{p\nu \log\log(np)} + \log\log(np)\right)\right\}.
\end{aligned}$$

考虑到由于 $\nu \geq \left(\sqrt{p}^{\,ns^*\log\log(np)} \right)^{\frac{2}{4\alpha^*+1}}$ 和 $s^* \geq p^{1/2+\delta}$，可立即得出 ν 关于 n 呈多项式增长，因此 $\nu \gtrsim \log\log n$。因此，对于某个普适正常数 κ'，我们有 $\sqrt{\cancel{rp}\cancel{\log\log(np)}} \geq \frac{1}{\kappa'}\log\log(np)$。然后根据切比雪夫不等式

$$\begin{aligned}
&P_{f^*}\left\{\chi^2_{\nu p}(n||f^*||_2^2) \leq \nu p + K_\eta\left(\sqrt{p\nu\log\log(np)} + \log\log(np)\right)\right\} \\
&\leq P_{f^*}\left\{\chi^2_{\nu p}(n||f^*||_2^2) \leq \nu p + K_\eta(1+\kappa')\left(\sqrt{p\nu\log\log(np)} + \log\log(np)\right)\right\} \\
&= P_{f^*}\left\{(n||f^*||_2^2 - K_\eta(1+\kappa')\sqrt{p\nu\log\log(np)}) \leq \nu p + n||f^*||_2^2 - \chi^2_{\nu p}(n||f^*||_2^2)\right\} \\
&\leq \frac{\text{Var}\left(\chi^2_{\nu p}(n||f^*||_2^2)\right)}{\left(n||f^*||_2^2 - K_\eta(1+\kappa')\sqrt{p\nu\log\log(np)}\right)^2} \\
&\leq \frac{2\nu p}{\left(n||f^*||_2^2 - K_\eta(1+\kappa')\sqrt{p\nu\log\log(np)}\right)^2} + \frac{4n||f^*||_2^2}{\left(n||f^*||_2^2 - K_\eta(1+\kappa')\sqrt{p\nu\log\log(np)}\right)^2}.
\end{aligned}$$

考虑到

$$\begin{aligned}
K_\eta(1+\kappa')\sqrt{p\nu\log\log(np)} &\leq K_\eta(1+\kappa')\sqrt{2\nu^* p\log\log(np)} \\
&\leq K_\eta(1+\kappa')\sqrt{2}ns^*\left(\frac{ns^*}{\sqrt{p\log\log(np)}}\right)^{-\frac{4\alpha^*}{4\alpha^*+1}} \\
&= \sqrt{2}K_\eta(1+\kappa')n\tau^2_{\text{dense}}(p, s^*, n, \alpha^*) \\
&\leq \frac{\sqrt{2}K_\eta(1+\kappa')}{C^2}n||f^*||_2^2.
\end{aligned}$$

因此，取 C_η 充分大，我们有

$$\begin{aligned}
&P_{f^*}\left\{\max_{\tilde{\nu}\in\mathcal{V}} \varphi_{\tilde{\nu}} = 0\right\} \\
&\leq \frac{2\nu p}{\left(n||f^*||_2^2 - K_\eta(1+\kappa')\sqrt{p\nu\log\log(np)}\right)^2} + \frac{4n||f^*||_2^2}{\left(n||f^*||_2^2 - K_\eta(1+\kappa')\sqrt{p\nu\log\log(np)}\right)^2} \\
&\leq \frac{2}{\left(\frac{C^2}{\sqrt{2}} - K_\eta(1+\kappa')\right)^2\log\log(np)} + \frac{4}{(1-\frac{\sqrt{2}K_\eta(1+\kappa')}{C^2})n||f^*||_2^2} \\
&\leq \frac{2}{\left(\frac{C^2}{\sqrt{2}} - K_\eta(1+\kappa')\right)^2\log\log(np)} + \frac{4}{(1-\frac{\sqrt{2}K_\eta(1+\kappa')}{C^2})\frac{C^2}{\sqrt{2}}} \\
&\leq \frac{\eta}{2}.
\end{aligned}$$

因此，第二类错误得到了适当的界定。由于 f^* 是任意的，我们证明了

$$P_0\left\{\max_{\nu\in\mathcal{V}}\varphi_\nu=1\right\}+\max_{s\geq p^{1/2+\delta}}\sup_{\alpha\in[\alpha_0,\alpha_1]}\sup_{\substack{f\in\mathcal{F}(s,\alpha),\\ ||f||_2\geq C\tau_{\mathrm{dense}}(p,s,n,\alpha)}}P_f\left\{\max_{\nu\in\mathcal{V}}\varphi_\nu=0\right\}\leq\eta$$

如所需。

□

参考文献

[1] E. Arias-Castro、E. J. Cand`es 和 Y. Plan。《稀疏备择下的全局检验：方差分析、多重比较与高阶批评》。《统计年刊》39.5（2011），第 2533–2556 页。[2] Y. Baraud。《信号检测的非渐近极小极大速率》。《伯努利》8.5（2002），第 577–606 页。[3] S. Bhattacharya、J. Fan 和 D. Mukherjee。《用于维数不断增长的非参数交互模型的深度神经网络》。《统计年刊》（即将出版）。[4] L. Birg´e 和 P. Massart。《高斯模型选择》。《欧洲数学会杂志》3.3（2001），第 203–268 页。[5] L. D. Brown 和 M. G. Low。《非参数回归与白噪声的渐近等价》。《统计年刊》24.6（1996），第 2384–2398 页。[6] M. V. Burnasev。《高斯白噪声背景下不完全已知信号的极小极大检测》。《概率论及其应用》24.1（1979），第 106–118 页。[7] A. Carpentier 和 N. Verzelen。《高斯向量模型中稀疏性的自适应估计》。《统计年刊》47.1（2019），第 93–126 页。[8] J. Chhor、R. Mukherjee 和 S. Sen。《异方差高斯序列模型中的稀疏信号检测：精确极小极大速率》。《伯努利》30.3（2024），第 2127–2153 页。[9] O. Collier、L. Comminges 和 A. B. Tsybakov。《稀疏正态均值的非光滑泛函估计》。《伯努利》26.3（2020），第 1989–2020 页。[10] O. Collier、L. Comminges 和 A. B. Tsybakov。《稀疏类上线性泛函和二次泛函的极小极大估计》。《统计年刊》45.3（2017），第 923–958 页。[11] O. Collier、L. Comminges、A. B. Tsybakov 和 N. Verzelen。《稀疏条件下线性泛函的最优自适应估计》。《统计年刊》46.6A（2018），第 3130–3150 页。[12] L. Comminges、O. Collier、M. Ndaoud 和 A. B. Tsybakov。《稀疏向量模型中的自适应稳健估计》。《统计年刊》49.3（2021），第 1347–1377 页。[13] N. Deb、R. Mukherjee、S. Mukherjee 和 M. Yuan。《伊辛模型中结构化信号的检测》。《应用概率年刊》34.1A（2024），第 1–45 页。[14] D. Donoho 和 J. Jin。《用于检测稀疏异质混合分布的高阶批评》。《统计年刊》32.3（2004），第 962–994 页。[15] M. S. Ermakov。《高斯白噪声中信号的极小极大检测》。《概率论及其应用》35.4（1990），第 704–715 页。[16] C. Gao、F. Han 和 C.-H. Zhang。《保序分段常数信号的估计》。《统计年刊》48.2（2020），第 629–654 页。

[17] G. Gayraud 和 Y. Ingster。《稀疏加性函数检测》。《电子统计学杂志》6（2012），第 1409–1448 页。[18] P. Hall 和 J. Jin。《利用创新高阶检验检测相关噪声中的稀疏信号》。《统计年刊》38.3（2010），第 1686–1732 页。[19] Y. Han、J. Jiao 和 R. Mukherjee。《高斯白噪声模型中 L_r-范数的估计》。《概率论与相关领域》177.3-4（2020），第 1243–1294 页。[20] T. Hastie 和 R. Tibshirani。《广义加性模型》。《统计科学》1.3（1986），第 297–318 页。[21] Y. I. Ingster。《非参数假设极小极大检验的自适应检验》。《概率论与数理统计》，第 I 卷（维尔纽斯，1989）。Mokslas，维尔纽斯，1990，第 539–549 页。[22] Y. I. Ingster。《非参数假设的渐近极小极大检验》。《概率论与数理统计》，第 I 卷（维尔纽斯，1985）。VNU Sci. Press，乌得勒支，1987，第 553–574 页。[23] Y. I. Ingster。《白高斯噪声中信号的极小极大非参数检测》。《信息传输问题》18.2（1982），第 61–73 页。[24] Y. I. Ingster 和 I. A. Suslina。《高斯模型下的非参数拟合优度检验》。第 169 卷。《统计学讲义》。Springer-Verlag，纽约，2003。[25] Y. Ingster 和 O. Lepski。《多通道非参数信号检测》。《数学方法与统计学》12.3（2003），第 247–275 页。[26] Y. I. Ingster、A. B. Tsybakov 和 N. Verzelen。《稀疏回归中的检测边界》。《电子统计学杂志》4（2010），第 1476–1526 页。[27] I. M. Johnstone。《高斯估计：序列模型与小波模型》。[28] I. M. Johnstone。《卡方预言不等式》。《概率论与统计学前沿》（莱顿，1999）。第 36 卷。统计学会研究所讲义专著系列。统计学会研究所，俄亥俄州比奇伍德，2001，第 399–418 页。[29] V. Koltchinskii 和 M. Yuan。《多核学习中的稀疏性》。《统计年刊》38.6（2010），第 3660–3695 页。[30] S. Kotekal 和 C. Gao。《相关条件下稀疏信号检测的极小极大速率》。《信息与推断》12.4（2023），论文编号 iaad044，共 97 页。[31] B. Laurent 和 P. Massart。《通过模型选择自适应估计二次泛函》。《统计年刊》28.5（2000），第 1302–1338 页。[32] O. Lepski、A. Nemirovski 和 V. Spokoiny。《回归函数 L_r 范数的估计》。《概率论与相关领域》113.2（1999），第 221–253 页。[33] Y. Lin 和 H. H. Zhang。《多元非参数回归中的成分选择与平滑》。《统计年刊》34.5（2006），第 2272–2297 页。[34] H. Liu、C. Gao 和 R. J. Samworth。《稀疏高维变点检测中的极小极大速率》。《统计年刊》49.2（2021），第 1081–1112 页。[35] L. Meier、S. van de Geer 和 P. B¨uhlmann。《高维加性建模》。《统计年刊》37.6B（2009），第 3779–3821 页。[36] R. Mukherjee 和 G. Ray。《伊辛模型中参数的检验》。《亨利·庞加莱研究所概率与统计年刊》58.1（2022），第 164–187 页。

[37] R. Mukherjee 和 S. Sen。《稀疏检验的极小极大指数》。arXiv:2003.00570[数学，统计学]。2020 年。[38] G. Raskutti、M. J. Wainwright 和 B. Yu。《通过凸规划得到核函数类上稀疏加性模型的极小极大最优收敛率》。机器学习研究杂志 13（2012），第 389–427 页。[39] P. Ravikumar、J. Lafferty、H. Liu 和 L. Wasserman。《稀疏加性模型》。英国皇家统计学会杂志B辑：统计方法论 71.5（2009），第 1009–1030 页。[40] M. Reiß。《具有多元随机设计的非参数回归的渐近等价》。《统计年刊》36.4（2008），第 1957–1982 页。[41] V. G. Spokoiny。《使用小波的自适应假设检验》。《统计年刊》24.6（1996），第 2477–2498 页。[42] N. M. Temme。《不完全伽马函数的渐近展开》。SIAM 数学分析杂志 10.4（1979），第 757–766 页。[43] N. M. Temme。《不完全伽马函数和不完全贝塔函数的均匀渐近展开》。数学计算 29.132（1975），第 1109–1114 页。[44] A. B. Tsybakov。《索об列夫类上函数的逐点和上确界范数尖锐自适应估计》。《统计年刊》26.6（1998），第 2420–2469 页。[45] A. B. Tsybakov。《非参数估计导论》。施普林格统计学丛书。Springer，纽约，2009 年。[46] R. Vershynin。《高维概率》。第 47 卷。剑桥统计与概率数学丛书。剑桥大学出版社，剑桥，2018 年。[47] M. J. Wainwright。《高维统计学》。第 48 卷。剑桥统计与概率数学丛书。剑桥大学出版社，剑桥，2019 年。[48] Y. Wei 和 M. J. Wainwright。《椭球中检验的局部几何：通过局部化柯尔莫哥洛夫宽度实现紧确控制》。IEEE信息论汇刊 66.8（2020），第 5110–5129 页。[49] Y. Yang 和 S. T. Tokdar。《高维空间中的极小极大最优非参数回归》。《统计年刊》43.2（2015），第 652–674 页。[50] M. Yuan 和 D.-X. Zhou。《高维加性模型中的极小极大最优估计收敛率》。《统计年刊》44.6（2016），第 2564–2593 页。[51] A. R. Zhang 和 Y. Zhou。《随机变量非渐近且尖锐的下尾界》。Stat 9（2020），e314，11 页。

硬阈值化

本节汇总随机变量（ $||Z||^2 - \alpha_t \ d \ \mathbb{I}^{\{||Z||^2 \geq d + t^2\}} \quad Z \sim （）$ ）对于 $N \ \theta \ I_d \quad \alpha_t \ d \quad 13 \quad t^2 \gtrsim d \quad t^2 \lesssim d$ （，）以及（）的相关结果，这些变量定义于（）。我们分别考察尾部和主体部分。

证明思路与 [34] 中采用的方法类似。不过，他们只考虑 $d = 1$，而我们需要考虑一般的 d 情形。因此，需要进行更为细致的分析。

A.1 尾部

引理 8（矩母函数）。 *Suppose $\widetilde{c}$ is a positive universal constant. There exist universal positive constants $D, \widetilde{C}, C$, and c such that if $d \geq D, t^2 \geq \widetilde{c}d$, and $0 < \lambda < \frac{1}{2C}$, then we have*

$$E\left(e^{\lambda Y}\right) \leq \exp\left(Ct^4\lambda^2 e^{-ct^2}\right)$$

where $Y = \left(||Z||^2 - \alpha_t(d)\right) \mathbb{I}_{\{||Z||^2 \geq d+t^2\}}$, $Z \sim N(0, I_d)$ and where $\alpha_t(d)$ is given by (13).

Proof. 我们沿用引理 18 的证明中所采用的论证思路，见 [34]。通用正常数 D 将在证明后文中选定，因此目前令 $d \geq D$ 且$Z \sim N(0, I_d)$。注意，我们还会令 D 足够大，以使引理21 适用。由于 $E(Y) = 0$，我们有

$$E(e^{\lambda Y}) = E(1 + \lambda Y + (e^{\lambda Y} - 1 - \lambda Y)) = 1 + E(e^{\lambda Y} - 1 - \lambda Y).$$

考虑对任意 $y \in \mathbb{R}$，我们有

$$e^y - 1 - y \leq \begin{cases} (-y) \wedge y^2 & \text{if } y < 0, \\ y^2 & \text{if } 0 \leq y \leq 1, \\ e^y & \text{if } y > 1. \end{cases}$$

因此，

$$E(e^{\lambda Y} - 1 - \lambda Y) \leq \lambda^2 E(Y^2 \mathbb{I}_{\{Y<0\}}) + \lambda^2 E(Y^2 \mathbb{I}_{\{0 \leq Y \leq 1/\lambda\}}) + E(e^{\lambda Y} \mathbb{I}_{\{Y > 1/\lambda\}}). \tag{49}$$

将分别对每一项进行界定。对于第一项，注意引理22 断言$\alpha_t(d) \leq d + C^*(t^2 \vee d)$，其中为通用常数。另请注意 $\alpha_t(d) \geq d + t^2$。因此，

$$\begin{aligned} \lambda^2 E(Y^2 \mathbb{I}_{\{Y<0\}}) &= \lambda^2 E((||Z||^2 - \alpha_t(d))^2 \mathbb{I}_{\{\alpha_t(d) > ||Z||^2 \geq d+t^2\}}) \\ &\leq \lambda^2 (\alpha_t(d) - d - t^2)^2 P\left\{\alpha_t(d) > ||Z||^2 \geq d + t^2\right\} \\ &\leq \lambda^2 (d + C^*(t^2 \vee d) - d - t^2)^2 P\left\{||Z||^2 \geq d + t^2\right\} \\ &\leq 2C^*\lambda^2 (d^2 \vee t^4) \exp\left(-c \min\left(\frac{t^4}{d}, t^2\right)\right) \\ &= 2C^*\lambda^2 t^4 \exp\left(-ct^2\right) \end{aligned}$$

其中 c 是一个普适正常数，而 C^* 仍是一个通用常数，但其取值可能因行而异。注意，在上面的推导式中，我们还使用了推论 1 以及 $t^2 \geq \widetilde{c}d$。综上，我们已经证明

$$\lambda^2 E(Y^2 \mathbb{I}_{\{Y<0\}}) \leq 2C^*\lambda^2 t^4 \exp\left(-ct^2\right). \tag{50}$$

现在我们对（49）中的第二项进行界定。令 f_d 表示 χ^2_d分布的概率密度函数。我们有

$$\begin{aligned} \lambda^2 E(Y^2 \mathbb{I}_{\{0<Y<1/\lambda\}}) &\leq \lambda^2 E(Y^2 \mathbb{I}_{\{0<Y\}}) \\ &= \lambda^2 E((||Z||^2 - \alpha_t(d))^2 \mathbb{I}_{\{||Z||^2 > \alpha_t(d)\}}) \\ &= \lambda^2 \int_{\alpha_t(d)}^{\infty} (z - \alpha_t(d))^2 f_d(z)\, dz \\ &\leq 2\lambda^2 \int_{\alpha_t(d)}^{\infty} \left(z^2 + \alpha_t(d)^2\right) f_d(z)\, dz. \end{aligned}$$

应用引理 18 可得

$$\begin{aligned}\int_{\alpha_t(d)}^{\infty} z^2 f_d(z)\,dz &= d(d+2)P\left\{\chi^2_{d+4} \geq \alpha_t(d)\right\}\\ &\leq 3d^2 P\left\{\chi^2_{d+4} \geq d+4+(\alpha_t(d)-d-4)\right\}\\ &\leq 3d^2 P\left\{\chi^2_{d+4} \geq d+4+(t^2-4)\right\}\\ &\leq 6d^2 \exp\left(-c\min\left(\frac{(t^2-4)^2}{d+4},(t^2-4)\right)\right)\\ &\leq 6d^2 \exp\left(-c\min\left(\frac{t^4}{20d},\frac{t^2}{2}\right)\right)\\ &\leq 6d^2 \exp\left(-c\min\left(\frac{t^4}{d},t^2\right)\right)\\ &= 6d^2\exp\left(-ct^2\right)\end{aligned}$$

其中 c 仍为一个通用常数，但其值可能随行变化。注意，我们使用了 $\alpha_t(d) \geq d+t^2$、推论 1以及 $t^2 \geq \tilde{c}d$。同样，考虑

$$\begin{aligned}\int_{\alpha_t(d)}^{\infty} \alpha_t(d)^2 f_d(z)\,dz &= \alpha_t(d)^2 P\{\chi^2_d \geq \alpha_t(d)\}\\ &\leq \alpha_t(d)^2 P\{\chi^2_d \geq d+t^2\}\\ &\leq 2(d+C^*t^2)^2\exp\left(-c\min\left(\frac{t^4}{d},t^2\right)\right)\\ &\leq C^*t^4\exp\left(-c\min\left(\frac{t^4}{d},t^2\right)\right)\\ &= C^*t^4\exp\left(-ct^2\right)\end{aligned}$$

其中我们使用了 $(a+b)^2 \leq 2a^2+2b^2$、 $t^2 \geq \tilde{c}d$、引理 22和推论 1。再次， C^* 和 c均为通用常数，但其值可能随行变化。因此，我们已证明

$$\lambda^2 E(Y^2 \mathbb{I}_{\{Y<0\}}) \leq C^*\lambda^2 t^4 \exp\left(-ct^2\right). \tag{51}$$

现在我们对 (49) 中的最后一项进行界定。考虑到对于 $0 \leq \lambda < \frac{1}{2}$，我们有

$$E(e^{\lambda Y}\mathbb{I}_{\{Y>1/\lambda\}}) = \int_{\alpha_t(d)+1/\lambda}^{\infty} e^{\lambda(z-\alpha_t(d))} f_d(z)\,dz = e^{-\lambda\alpha_t(d)}\int_{\alpha_t(d)+1/\lambda}^{\infty}\frac{1}{2^{d/2}\Gamma(d/2)}e^{-\left(\frac{1}{2}-\lambda\right)z}z^{\frac{d}{2}-1}\,dz.$$

令 $u=\left(\frac{1}{2}-\lambda\right)z$。于是 $du=\left(\frac{1}{2}-\lambda\right)dz$，因此

$$\begin{aligned}e^{-\lambda\alpha_t(d)}\int_{\alpha_t(d)+1/\lambda}^{\infty}\frac{1}{2^{d/2}\Gamma(d/2)}e^{-\left(\frac{1}{2}-\lambda\right)z}z^{\frac{d}{2}-1}\,dz &= \frac{e^{-\lambda\alpha_t(d)}}{2^{d/2}\Gamma(d/2)}\int_{(\alpha_t(d)+1/\lambda)(1/2-\lambda)}^{\infty} e^{-u}u^{\frac{d}{2}-1}\left(\frac{1}{2}-\lambda\right)^{-\frac{d}{2}}\,du\\ &= \frac{e^{-\lambda\alpha_t(d)}}{(1-2\lambda)^{d/2}}\cdot\frac{1}{\Gamma\left(\frac{d}{2}\right)}\int_{(\alpha_t(d)+1/\lambda)(1/2-\lambda)}^{\infty} e^{-u}u^{\frac{d}{2}-1}\,du\\ &= \frac{e^{-\lambda\alpha_t(d)}}{(1-2\lambda)^{d/2}}\cdot\frac{\Gamma\left(\frac{d}{2},\left(\frac{1}{\lambda}+\alpha_t(d)\right)\left(\frac{1}{2}-\lambda\right)\right)}{\Gamma\left(\frac{d}{2}\right)}\end{aligned}$$

其中 $\Gamma(s,x)=\int_x^{\infty}e^{-t}t^{s-1}\,dt$。总之，我们已证明

$$E(e^{\lambda Y}\mathbb{I}_{\{Y>1/\lambda\}}) = \frac{e^{-\lambda\alpha_t(d)}}{(1-2\lambda)^{d/2}}\cdot\frac{\Gamma\left(\frac{d}{2},\left(\frac{1}{\lambda}+\alpha_t(d)\right)\left(\frac{1}{2}-\lambda\right)\right)}{\Gamma\left(\frac{d}{2}\right)}. \tag{52}$$

为了继续，我们希望应用推论 2，但必须先验证相关条件。令 $a=\frac{d}{2}$ 并令 $\eta=\sqrt{2(\mu-\log(1+\mu))}$，其中 $\mu=\left(\frac{1}{\lambda}+\alpha_t(d)\right)\left(\frac{1}{2}-\lambda\right)_a - 1$。注意，我们可以将 D取为足够大的通用常数，以确保 a 足够大。由于 $\tilde{c}d \leq t^2$，由引理22 可知， $\alpha_t(d) \leq \widetilde{C}t^2$，其中某个正的通用常数为 $\widetilde{C}$。不失一般性，我们可以取

$\widetilde{C} \geq \frac{3}{2}$。让我们将注意力限制在 $\lambda < \frac{1}{2\widetilde{C}}$ 上。注意到

$$\mu = \frac{\frac{1}{\lambda} + \alpha_t(d) - 2 - 2\lambda\alpha_t(d) - d}{d} \geq \frac{(2\widetilde{C} - 2) + t^2 - \frac{2}{2\widetilde{C}}(\widetilde{C}t^2)}{d} \geq \frac{2(\widetilde{C} - 1)}{d} = \frac{C^{**}}{d}$$

其中我们定义了 $C^{**} = 2(\widetilde{C} - 1)$ 。由于我们已证明 $\mu > 0$，因此可以应用推论2。根据推论2，我们有

$$\frac{\Gamma\left(\frac{d}{2}, \left(\frac{1}{\lambda} + \alpha_t(d)\right)\left(\frac{1}{2} - \lambda\right)\right)}{\Gamma\left(\frac{d}{2}\right)} \leq \left(1 - \Phi\left(\eta\sqrt{a}\right)\right) + \frac{C^*}{\sqrt{d}} \exp\left(-\frac{a\eta^2}{2}\right).$$

这里， Φ 表示标准正态分布的累积分布函数， C^* 是一个通用常数。根据高斯尾界 $1 - \Phi(x) \leq {}^{\varphi(x)\ x}$，对于满足 $\varphi = \Phi'$的 $x > 0$，我们有

$$\frac{\Gamma\left(\frac{d}{2}, \left(\frac{1}{\lambda} + \alpha_t(d)\right)\left(\frac{1}{2} - \lambda\right)\right)}{\Gamma\left(\frac{d}{2}\right)} \leq C^* \exp\left(-\frac{a\eta^2}{2}\right)\left(\frac{1}{\eta\sqrt{a}} + \frac{1}{\sqrt{d}}\right)$$

其中， C^* 的值虽随行而变，但始终是一个通用常数。为继续计算，我们需要从下方对$\eta\sqrt{a}$进行界定。注意，我们可以取 $D \geq C^{**}$。对于 $\lambda < \frac{1}{2\widetilde{C}}$，由 $\frac{C^{**}}{d} \leq 1$ 可得

$$\begin{aligned}
\eta\sqrt{a} &= \sqrt{a}\sqrt{2(\mu - \log(1+\mu))} \\
&\geq \sqrt{a}\sqrt{2\left(\frac{C^{**}}{d} - \log\left(1 + \frac{C^{**}}{d}\right)\right)} \\
&\geq \sqrt{a}\sqrt{2 \cdot \left(\frac{(C^{**})^2}{2d} - \frac{(C^{**})^3}{3d^3}\right)} \\
&= \sqrt{\frac{(C^{**})^2}{2d} - \frac{(C^{**})^3}{3d^2}} \\
&= \frac{C^{**}}{\sqrt{d}}\sqrt{\frac{1}{2} - \frac{C^{**}}{3d}} \\
&\geq \frac{C^{**}}{\sqrt{6d}}.
\end{aligned}$$

因此，我们得到如下界

$$\frac{\Gamma\left(\frac{d}{2}, \left(\frac{1}{\lambda} + \alpha_t(d)\right)\left(\frac{1}{2} - \lambda\right)\right)}{\Gamma\left(\frac{d}{2}\right)} \leq C^*\sqrt{d}\exp\left(-\frac{a\eta^2}{2}\right)$$

其中， C^* 的值虽有变化，但始终是一个通用常数。现在我们来考察项$e^{-a\eta^2/2}$。注意到

$$\begin{aligned}
\exp\left(-\frac{a\eta^2}{2}\right) &= \exp\left(-a\left(\mu - \log\left(1+\mu\right)\right)\right) \\
&= \exp\left(-a\mu\right)\left(\frac{\left(\frac{1}{\lambda} + \alpha_t(d)\right)(1 - 2\lambda)}{d}\right)^{d/2} \\
&= \exp\left(\frac{d}{2} - \frac{1}{2\lambda} + 1 - \frac{\alpha_t(d)}{2} + \lambda\alpha_t(d)\right)\left(\frac{\left(\frac{1}{\lambda} + \alpha_t(d)\right)(1 - 2\lambda)}{d}\right)^{d/2}.
\end{aligned}$$

因此，令 C^* 的值逐行变化，但始终为通用常数，

we

由（52）可得

$$
\begin{aligned}
E\left(e^{\lambda Y}\mathbb{I}_{\{Y>1/\lambda\}}\right) &\le C^*\sqrt{d}\frac{e^{-\lambda\alpha_t(d)}}{(1-2\lambda)^{d/2}}\exp\left(\frac{d}{2}-\frac{1}{2\lambda}+1-\frac{\alpha_t(d)}{2}+\lambda\alpha_t(d)\right)\left(\frac{\left(\frac{1}{\lambda}+\alpha_t(d)\right)(1-2\lambda)}{d}\right)^{d/2}\\
&\le C^*\sqrt{d}e^{-\left(\frac{1}{2\lambda}+\frac{\alpha_t(d)-d}{2}\right)}\left(1+\frac{\frac{1}{\lambda}+\alpha_t(d)-d}{d}\right)^{d/2}\\
&= C^*\sqrt{d}\exp\left(-\left(\frac{1}{2\lambda}+\frac{\alpha_t(d)-d}{2}\right)+\frac{d}{2}\log\left(1+\frac{\frac{1}{\lambda}+\alpha_t(d)-d}{d}\right)\right)\\
&= C^*\sqrt{d}\exp\left(-\frac{d}{2}\left(\left(\frac{\frac{1}{\lambda}+\alpha_t(d)-d}{d}\right)-\log\left(1+\frac{\frac{1}{\lambda}+\alpha_t(d)-d}{d}\right)\right)\right)\\
&\le C^*\sqrt{d}\exp\left(-\frac{d}{2}\cdot\frac{\frac{1}{\lambda}+\alpha_t(d)-d}{c^{**}d}\right)\\
&\le C^*\sqrt{d}\exp\left(-\frac{1}{2c^{**}\lambda}-\frac{t^2}{2c^{**}}\right)
\end{aligned}
$$

对于某个普适正常数 c^{**}。我们使用了对于 $u\ge 1$，有 $u-\log(1+u)\gtrsim u$。注意，由于$\frac{1}{\lambda}+\alpha_t(d)-d\gtrsim\frac{t^2}{d}\gtrsim 1$），~~我们可以这~~样处理。由于对于所有 $u\in\mathbb{R}$ 都有 $e^{-\frac{1}{2c^{**}u}}\lesssim u^2$，因此可得

$$
E\left(e^{\lambda Y}\mathbb{I}_{\{Y>1/\lambda\}}\right)\le C^*\sqrt{d}\lambda^2\exp\left(-\frac{t^2}{2c^{**}}\right) \tag{53}
$$

其中 C^* 的取值再次发生了变化，但仍为通用常数。将我们的界（50）、（51）、（53）代入（49）可得

$$
E(e^{\lambda Y})\le 1+Ct^4\lambda^2e^{-ct^2}\le\exp\left(Ct^4\lambda^2e^{-ct^2}\right)
$$

对于 $\lambda<\frac{1}{2C}$，其中 C、$c>0$ 为通用常数。证明完毕。 □

引理 9。 *Let $Z\sim N(\theta,\ I_d)$. Suppose $\widetilde{c}$ is a universal positive constant. There exists a universal constant C such that for every $t^2\ge\widetilde{c}d$, we have*

$$
E\left\{(||Z||^2-\alpha_t(d))\mathbb{I}_{\{||Z||^2\ge d+t^2\}}\right\}\begin{cases}=0 & \text{if } \theta=0,\\ \ge 0 & \text{if } ||\theta||^2<Ct^2,\\ \ge ||\theta||^2/2 & \text{if } ||\theta||^2\ge Ct^2.\end{cases}
$$

Here, $\alpha_t(d)$ is given by (13).

Proof. 我们将在证明结束时确定 C。先考虑 $\theta=0$ 的情形。根据 $\alpha_t(d)$ 的定义，有 $E\{(||Z||^2-\alpha_t(d))\mathbb{I}_{\{||Z||^2\ge d+t^2\}}\}=0$，因此可得所需结果。第二种情形成立，是因为期望中的表达式关于 $||\theta||$ 随机递增。接下来考虑最后一种情形，设 $||\theta||^2\ge Ct^2$。由于 $\alpha_t(d)\ge d+t^2$，我们有

$$
E\left\{(||Z||^2-\alpha_t(d))\mathbb{I}_{\{||Z||^2<d+t^2\}}\right\}\le 0
$$

因此，

$$
\begin{aligned}
E\left\{(||Z||^2-\alpha_t(d))\mathbb{I}_{\{||Z||^2\ge d+t^2\}}\right\} &= E\left\{(||Z||^2-\alpha_t(d))(1-\mathbb{I}_{\{||Z||^2<d+t^2\}})\right\}\\
&= d+||\theta||^2-\alpha_t(d)-E\left\{(||Z||^2-\alpha_t(d))\mathbb{I}_{\{||Z||^2<d+t^2\}}\right\}\\
&\ge d+||\theta||^2-\alpha_t(d)
\end{aligned}
$$

根据引理22 和 $t^2\gtrsim d$，存在一个普适正常数 C^* 使得 $\alpha_t(d)\le d+C^*t^2$。因此，

$$
E\left\{(||Z||^2-\alpha_t(d))\mathbb{I}_{\{||Z||^2\ge d+t^2\}}\right\}\ge||\theta||^2-C^*t^2\ge||\theta||^2\left(1-\frac{C^*}{C}\right).
$$

选取 $C = \frac{C^*}{2}$ 即可得到所需结果。 □

引理 10. *Let $Z \sim N(\theta, I_d)$. If $t^2 \geq \widetilde{c}d$ for a universal positive constant $\widetilde{c}$, then*

$$\mathrm{Var}\left(\left(||Z||^2 - \alpha_t(d)\right)_{\mathbb{I}\{||Z||^2 \geq d+t^2\}}\right) \lesssim \begin{cases} t^4 \exp\left(-c^* \min\left(\frac{t^4}{d}, t^2\right)\right) & \text{if } \theta = 0, \\ t^4 & \text{if } 0 < ||\theta|| < 2t, \\ ||\theta||^2 & \text{if } ||\theta|| > 2t. \end{cases}$$

Here, $\alpha_t(d)$ is given by (13) and c^* is a universal positive constant.

Proof. 引理的证明14 可通过作如下修改而重复：调用引理 22 而不是引理21。 □

引理 11. *Suppose $\widetilde{c}$ is a universal positive constant. There exist universal positive constants D and C^* such that if $d \geq D$, $t^2 \geq \widetilde{c}d$, and $x > 0$, then*

$$P\left\{\sum_{i=1}^{n}\left(||Z_i||^2 - \alpha_t(d)\right)_{\mathbb{I}\{||Z||^2 \geq d+t^2\}} \geq C^*\left(\sqrt{xnt^4e^{-ct^2}} + x\right)\right\} \leq e^{-x}$$

where $Z_1, ..., Z_n \overset{iid}{\sim} N(0, I_d)$ and $\alpha_t(d)$ is given by (13). Here, c is a universal positive constant.

Proof. 设 D 为引理 8中的常数。我们使用切尔诺夫方法得到所需界。设 Y 如引理 8 所述，并设 $\widetilde{C}$ 为引理 8中的通用常数。对于任意 $u > 0$，由引理 8

$$\begin{aligned} P\left\{\sum_{j=1}^{n}\left(||Z_j||^2 - \alpha_t(d)\right)_{\mathbb{I}\{||Z_j||^2 \geq d+t^2\}} > u\right\} &\leq \inf_{\lambda < \frac{1}{2\widetilde{C}}} e^{-\lambda u}\left(E(e^{\lambda Y})\right)^n \\ &\leq \inf_{\lambda < \frac{1}{2\widetilde{C}}} \exp\left(-\lambda u + Cnt^4\lambda^2 e^{-ct^2}\right) \\ &= \exp\left(-C\left(\frac{u^2}{nt^4e^{-ct^2}} \wedge u\right)\right) \end{aligned}$$

其中， $c > 0$ 是一个通用常数， $C > 0$ 也是一个通用常数，其值可能逐行变化，但始终是通用常数。适当选取 $u = \frac{1}{2}\left(\sqrt{\frac{xnt^4e^{-ct^2}}{C}} + \frac{x}{C}\right)$ 和 C^* 即可得到所需结果。证明完毕。 □

A.2 主体部分

引理 12. *Let L^* be the universal positive constant from Lemma 21. There exist universal constants C^*, $C^{**}, C, c > 0$ such that if $d \geq C^{**}$ and $1 \leq \beta \leq L^*\sqrt{d}$, then*

$$E(e^{\lambda Y}) \leq \exp\left(Cd\beta^2\lambda^2 e^{-c\beta^2}\right)$$

for $\lambda < \frac{\beta}{2(\beta C^* + \sqrt{d})}$. Here, $Y = \left(||Z||^2 - \alpha_t(d)\right)_{\mathbb{I}\{||Z||^2 \geq d+t^2\}}$ with $t^2 = \beta\sqrt{d}$ and $Z \sim N(0, I_d)$. Also, $\alpha_t(d)$ is given by (13).

Proof. 我们给出的论证将类似于引理 8 的证明，即我们将分别界定（49）中的每一项，并代入该方程

$$E(e^{\lambda Y}) = 1 + E\left(e^{\lambda Y} - 1 - \lambda Y\right).$$

我们先看（49）右侧的第一项。根据引理18，对于某个普适正常数 α_t（d），有 $\leq d + C^*\beta\sqrt{d}$。注意，$C^*$因为显然有 $C^* \geq 1$ α_t（d） $\geq d + t^2 = d + \beta\sqrt{d}$。

因此，仿照引理8的证明可得

$$\begin{aligned}\lambda^2 E(Y^2 \mathbb{I}_{\{Y<0\}}) &\le \lambda^2 (d+t-\alpha_t(d))^2 P\{||Z||^2 \ge d+t^2\} \\ &\le C\lambda^2 d\beta^2 \exp\left(-c\min\left(\frac{t^4}{d}, t^2\right)\right) \\ &= C\lambda^2 d\beta^2 e^{-c\beta^2}\end{aligned}$$

其中 C， $c>0$ 是通用常数。我们使用推论 1 得到了最后一个不等式。现在我们对（49）中的第二项进行界定。令 f_d 表示 χ^2_d分布的概率密度函数，我们重复引理 8 证明中的步骤，继续计算，得到

$$\begin{aligned}\lambda^2 E(Y^2 \mathbb{I}_{\{0<Y<1/\lambda\}}) &\le \lambda^2 \int_{\alpha_t(d)}^{\infty} (z-\alpha_t(d))^2 f_d(z)\,dz \\ &\le \lambda^2 \int_{d+t^2}^{\infty} (z-\alpha_t(d))^2 f_d(z)\,dz \\ &= \lambda^2 \left(\int_{d+t^2}^{\infty} z^2 f_d(z)\,dz - 2\alpha_t(d)\int_{d+t^2}^{\infty} z f_d(z)\,dz + \alpha_t(d)^2 \int_{d+t^2}^{\infty} f_d(z)\,dz\right) \\ &= \lambda^2 P\{\chi^2_d \ge d+t^2\}\,\mathrm{Var}\left(||Z||^2 \,|\, ||Z||^2 \ge d+t^2\right) \\ &\le C\lambda^2 d\beta^2 \exp\left(-c\min\left(\frac{t^4}{d}, t^2\right)\right) \\ &= C\lambda^2 d\beta^2 e^{-c\beta^2}\end{aligned}$$

其中 C， $c>0$ 的取值虽已改变，但仍为通用常数。这里我们使用了引理 24。

现在我们对（49）中的最后一项进行界定。完全按照引理 8的证明，我们有

$$E(e^{\lambda Y} \mathbb{I}_{\{Y>1/\lambda\}}) = \frac{e^{-\lambda\alpha_t(d)}}{(1-2\lambda)^{d/2}} \cdot \frac{\Gamma\left(\frac{d}{2}, \left(\frac{1}{\lambda}+\alpha_t(d)\right)\left(\frac{1}{2}-\lambda\right)\right)}{\Gamma\left(\frac{d}{2}\right)}$$

对于 $0 \le \lambda < \frac{1}{2}$。回想一下， $\Gamma(s,x) = \int_x^{\infty} e^{-t} t^{s-1}\,dt$。

我们将注意力限制在 $\lambda < \frac{1}{2} - \frac{\beta}{(\sqrt{d}+C^*\beta)}$。我们希望应用推论 2，但必须验证其条件。令 $a = \frac{d}{2}$，并令 $\eta = \sqrt{2(\mu - \log(1+\mu))}$，其中 $\mu = \left(\frac{1}{\lambda}+\alpha_t(d)\right)\left(\frac{1}{2}-\lambda\right) - 1$。注意

$$\mu = \frac{\left(\frac{1}{\lambda}+\alpha_t(d)\right)(1-2\lambda) - d}{d} = \frac{\frac{1-2\lambda}{\lambda} + \alpha_t(d)(1-2\lambda) - d}{d}.$$

考虑 $\lambda < \frac{1}{2} - \frac{\beta}{(\sqrt{d}+C^*\beta)} = \frac{1}{2} - \frac{t^2}{(d+C^*\beta\sqrt{d})} \le \frac{\alpha_t}{\left(\frac{d}{2\alpha_t} - d\right)}$。因此， $\alpha_t(d)(1-2\lambda) - d \ge 0$ 于是

$$\mu \ge \frac{1-2\lambda}{\lambda d} \ge \frac{2}{\beta\sqrt{d}}.$$

注意，我们使用了 $C^* \ge 1$。由于 $\mu > 0$ 且 $d \ge C^{**}$ 是一个足够大的通用常数，因此我们可以应用推论 2。根据推论 2，我们有

$$\frac{\Gamma\left(\frac{d}{2}, \left(\frac{1}{\lambda}+\alpha_t(d)\right)\left(\frac{1}{2}-\lambda\right)\right)}{\Gamma\left(\frac{d}{2}\right)} \le \left(1-\Phi\left(\eta\sqrt{a}\right)\right) + \frac{C^\dagger}{\sqrt{d}} \exp\left(-\frac{a\eta^2}{2}\right).$$

这里， $C^\dagger$ 是一个正的通用常数。回忆一下， Φ 表示标准正态分布的累积分布函数。根据高斯尾界 $1-\Phi(x) \le \varphi(x)\,x$，对于 $x>0$，我们有

$$\frac{\Gamma\left(\frac{d}{2}, \left(\frac{1}{\lambda}+\alpha_t(d)\right)\left(\frac{1}{2}-\lambda\right)\right)}{\Gamma\left(\frac{d}{2}\right)} \le C^\dagger \exp\left(-\frac{a\eta^2}{2}\right)\left(\frac{1}{\eta\sqrt{a}} + \frac{1}{\sqrt{d}}\right) \tag{54}$$

其中 $C^\dagger$ 的值已发生变化，但仍是一个通用常数。为了继续进行 bound，

我们需要给出 $\eta\sqrt{a}$ 的下界。设 C^{**} 大于 4。由于 $a=\frac{d}{2}$，我们有

$$\begin{aligned}\eta\sqrt{a} &= \sqrt{a}\sqrt{2(\mu-\log(1+\mu))}\\ &\geq \sqrt{a}\sqrt{2\left(\frac{2}{\beta\sqrt{d}}+\log\left(1+\frac{2}{\beta\sqrt{d}}\right)\right)}\\ &\geq \sqrt{a}\sqrt{2\left(\frac{4}{2\beta^2 d}-\frac{8}{3\beta^3 d^{3/2}}\right)}\\ &\geq \frac{c^*}{\beta}\end{aligned}$$

其中 $c^*>0$ 是一个通用常数。由式（54）可得

$$\frac{\Gamma\left(\frac{d}{2},\left(\frac{1}{\lambda}+\alpha_t(d)\right)\left(\frac{1}{2}-\lambda\right)\right)}{\Gamma\left(\frac{d}{2}\right)} \leq C^{\dagger}\beta\exp\left(-\frac{a\eta^2}{2}\right)$$

其中 $C^{\dagger}$ 的取值虽已改变，但仍是一个通用常数。现在考察项 $e^{-\frac{a\eta^2}{2}}$。完全按照引理 8的证明进行推导，可得

$$\exp\left(-\frac{a\eta^2}{2}\right)=\exp\left(\frac{d}{2}-\frac{1}{2\lambda}+1-\frac{\alpha_t(d)}{2}+\lambda\alpha_t(d)\right)\left(\frac{\left(\frac{1}{\lambda}+\alpha_t(d)\right)(1-2\lambda)}{d}\right)^{d/2}$$

正如引理 8的证明中所述，由此可得界

$$E(e^{\lambda Y}\mathbb{I}_{Y>1/\lambda}) \leq C^*\beta\exp\left(C\lambda^2 d\beta^2 e^{-c\beta^2}\right)$$

对于 $\lambda<\frac{e^{-\beta}}{2(\sqrt{d}+C^*\beta)}$，即得所需结论。 □

引理 13。 *Let $Z\sim N(\theta,\ I_d)$. Suppose $1\leq\beta\leq L^*\sqrt{d}$ where L^* is the universal constant from Lemma 21. There exists a universal positive constant C such that if $t^2=\beta\sqrt{d}$, then we have*

$$E\left\{(||Z||^2-\alpha_t(d))\mathbb{I}_{\{||Z||^2\geq d+t^2\}}\right\}\begin{cases}=0 & \text{if } \theta=0,\\ \geq 0 & \text{if } ||\theta||^2<Ct^2,\\ \geq ||\theta||^2/2 & \text{if } ||\theta||^2\geq Ct^2\end{cases}$$

where $\alpha_t(d)$ is given by (13).

Proof. 我们将在证明的后面确定 C。对于情形 $\theta=0$ 和 $||\theta||^2\leq Ct^2$，证明与引理9的证明完全相同。这里，我们重点讨论最后一种情形，其中 $||\theta||^2\geq Ct^2$。由于 $\alpha_t(d)\geq d+t^2$，我们有

$$E\left\{(||Z||^2-\alpha_t(d))\mathbb{I}_{\{||Z||^2<d+t^2\}}\right\}\leq 0.$$

由于$t^2=\beta\sqrt{d}$，我们可以应用引理21 得到

$$\begin{aligned}E\left\{(||Z||^2-\alpha_t(d))\mathbb{I}_{\{||Z||^2\geq d+t^2\}}\right\} &= E\left\{(||Z||^2-\alpha_t(d))(1-\mathbb{I}_{\{||Z||^2<d+t^2\}})\right\}\\ &\geq d+||\theta||^2-\alpha_t(d)-E\left\{(||Z||^2-\alpha_t(d))\mathbb{I}_{\{||Z||^2<d+t^2\}}\right\}\\ &\geq ||\theta||^2-C^*\beta\sqrt{d}\\ &= ||\theta||^2\left(1-\frac{C^*\beta\sqrt{d}}{||\theta||^2}\right)\\ &= ||\theta||^2\left(1-\frac{C^*t^2}{||\theta||^2}\right)\\ &\geq ||\theta||^2\left(1-\frac{C^*}{C}\right).\end{aligned}$$

其中 C^* 是一个普适正常数。取 $C=\frac{C^*}{2}$ 即可完成证明。 □

引理 14。 *Let* $Z \sim N(\theta, I_d)$. *Suppose* $1 \le \beta \le L^*\sqrt{d}$ *where* L^* *is the universal constant from Lemma 21. Then there exists a universal positive constant* c^* *such that if* $t^2 = \beta\sqrt{d}$, *then*

$$\mathrm{Var}\left(\left(||Z||^2 - \alpha_t(d)\right)_{\mathbb{I}\{||Z||^2 \ge d+t^2\}}\right) \lesssim \begin{cases} t^4 \exp\left(-\frac{c^* t^4}{d}\right) & \text{if } \theta = 0, \\ d + ||\theta||^2 & \text{if } ||\theta|| \ge 2t, \\ t^4 & \text{if } 0 < ||\theta|| < 2t. \end{cases}$$

Here, $\alpha_t(d)$ *is given by (13).*

Proof. 令 f_d 和 F_d 分别表示 χ^2_d 分布的概率密度函数和累积分布函数。

情形 1：考虑满足 $\theta = 0$的第一种情形。于是

$$\begin{aligned} \mathrm{Var}((||Z||^2 - \alpha_t(d))_{\mathbb{I}\{||Z||^2 \ge d+t^2\}}) &\le E\left((||Z||^2 - \alpha_t(d))^2_{\mathbb{I}\{||Z||^2 \ge d+t^2\}}\right) \\ &= P\left\{||Z||^2 \ge d+t^2\right\} \cdot E\left((||Z||^2 - \alpha_t(d))^2 \,|\, ||Z||^2 \ge d+t^2\right) \\ &= \left(1 - F_d(d+t^2)\right) \mathrm{Var}\left(||Z||^2 \,|\, ||Z||^2 \ge d+t^2\right) \\ &\le C^* \left(1 - F_d(d+t^2)\right) d\beta^2 \end{aligned}$$

其中我们应用了 $\alpha_t(d)$的定义，并应用了引理24。这里， C^* 是一个普适正常数。根据推论1 可得该情形所需的结果。

现在转到另外两种情形。设 $\theta \ne 0$。为简化记号，令 $Y = (||Z||^2 - \alpha_t(d))\mathbb{I}^{\{||Z||^2 \ge d+t^2\}}$。注意

$$\begin{aligned} &\mathrm{Var}(Y) \\ &= E\left(\text{方差}(Y \,|\, _{\mathbb{I}\{||Z||^2 \ge d+t^2\}})\right) + \text{方差}\left(E(Y \,|\, _{\mathbb{I}\{||Z||^2 \ge d+t^2\}})\right) \le P\left\{||Z||^2 \ge d+t^2\right\}\text{方差} \\ &\left(||Z||^2 \,|\, ||Z||^2 \ge d+t^2\right) + P\left\{||Z||^2 \ge d+t^2\right\} \left(E(||Z||^2 - \alpha_t(d) \,|\, ||Z||^2 \ge d+t^2)\right)^2. \end{aligned} \tag{55}$$

注意，由于 $\theta \ne 0$（通过应用随机序）， $0 \le E(||Z||^2 - \alpha_t(d) \,|\, ||Z||^2 \ge d+t^2)$，因此我们可以先求条件期望的上界，再求该条件期望平方的上界。我们分别考察上面公式中的每一项。首先，考虑

$$\begin{aligned} P\left\{||Z||^2 \ge d+t^2\right\} \mathrm{Var}\left(||Z||^2 \,|\, ||Z||^2 \ge d+t^2\right) &\le E\left(\mathrm{Var}\left(||Z||^2 \,|\, _{\mathbb{I}\{||Z||^2 \ge d+t^2\}}\right)\right) \\ &\le \mathrm{Var}(||Z||^2) \\ &= 2d + 4||\theta||^2. \end{aligned} \tag{56}$$

现在考察第二项。注意，我们可以写成 $Z = g + \theta$，其中 $g \sim N(0, I_d)$。因此，

$$\begin{aligned} E(||Z||^2 - \alpha_t(d) \,|\, ||Z||^2 \ge d+t^2) &= E\left(||g+\theta||^2 - \alpha_t(d) \,|\, ||g+\theta||^2 \ge d+t^2\right) \\ &\le E(||g||^2 + 2\langle\theta, g\rangle + ||\theta||^2 - d - t^2 \,|\, ||g+\theta||^2 \ge d+t^2) \\ &= E(||g||^2 - d \,|\, ||g+\theta||^2 \ge d+t^2) + ||\theta||^2 - t^2 \end{aligned} \tag{57}$$

其中我们利用了 $\langle\theta、 g\rangle_{\mathbb{I}\{||g+\theta||^2 \ge d+t^2\}} \overset{d}{=} \langle\theta、 -g\rangle_{\mathbb{I}\{||-g+\theta||^2 \ge d+t^2\}}$。由此可得 $E(\langle\theta, g\rangle_{\mathbb{I}\{||g+\theta||^2 \ge d+t^2\}}) = 0$，进而推出 $E(\langle\theta, g\rangle \,|\, ||g+\theta||^2 \ge d+t^2) = 0$。注意，为了得到上述显示式中的第二行，我们还利用了 $\alpha_t(d) \ge d + t^2$。有了上述显示式后，我们现在考察剩下的两种情况。

情形 2：考虑情形 $||\theta|| \ge 2t$。注意到

$$E(||g||^2 - d \,|\, ||g+\theta||^2 \ge d+t^2) = \frac{E\left((||g||^2 - d)_{\mathbb{I}\{||g+\theta||^2 \ge d+t^2\}}\right)}{P\left\{||g+\theta||^2 \ge d+t^2\right\}}.$$

考察分母，由于 $||g+\theta||^2 \sim \chi_d^2(||\theta||^2) \overset{d}{=} \chi_{d-1}^2 + \chi_1^2(||\theta||^2)$，且右侧的两个 χ^2 随机变量相互独立，我们有

$$P\{||g+\theta||^2 \geq d+t^2\} \geq P\{\chi_{d-1}^2 \geq d-1\} P\{\chi_1^2(||\theta||^2) \geq 1+t^2\} \geq cP\{\chi_1^2(||\theta||^2) \geq 1+t^2\}$$

其中 c 是一个普适正常数。考察分子，注意到根据引理18 ，我们有

$$\begin{aligned} E\left((||g||^2-d)_{\mathbb{I}\{||g+\theta||^2 \geq d+t^2\}}\right) &\leq E((||g||^2-d)_{\mathbb{I}\{||g||^2 \geq d\}}) \\ &= \int_d^\infty (z-d) f_d(z)\, dz \\ &= d\int_d^\infty (f_{d+2}(z) - f_d(z))\, dz \\ &= -2d\int_d^\infty f'_{d+2}(z)\, dz \\ &= 2df_{d+2}(d). \end{aligned}$$

因此，由式（57），我们得到

$$E(||Z||^2 - \alpha_t(d) \,|\, ||Z||^2 \geq d+t^2) \leq \frac{2df_{d+2}(d)}{cP\{\chi_1^2(||\theta||^2) \geq 1+t^2\}} + ||\theta||^2 - t^2.$$

因此，我们有如下界

$$\mathrm{Var}\,(Y) \leq 2d + 4||\theta||^2 + P\{||Z||^2 \geq d+t^2\} \left(\frac{2df_{d+2}(d)}{cP\{\chi_1^2(||\theta||^2) \geq 1+t^2\}} + ||\theta||^2 - t^2\right)^2.$$

注意到

$$\frac{2df_{d+2}(d)}{c} = \frac{d}{c\Gamma\left(\frac{d}{2}+1\right)} \left(\frac{d}{2e}\right)^{d/2} \leq c\sqrt{d}$$

其中， c 的值在各个表达式中可以不同，但始终是一个普适正常数。最后一个不等式由斯特林公式得出。此外，由于 $||\theta|| \geq 2t$，存在一个普适正常数 c'，使得 $P\{\chi_1^2(||\theta||^2) \geq 1+t^2\} \geq 1/c'$。进一步，由切比雪夫不等式可得

$$P\{||Z||^2 < d+t^2\} = P\{||\theta||^2 - t^2 < d + ||\theta||^2 - ||Z||^2\} \leq \frac{2d+4||\theta||^2}{(||\theta||^2-t^2)^2}.$$

因此，

$$\begin{aligned} \mathrm{Var}(Y) &\leq 2d + 4||\theta||^2 + \frac{2d+4||\theta||^2}{(||\theta||^2-t^2)^2}(cc'\sqrt{d} + ||\theta||^2 - t^2)^2 \\ &\asymp d + ||\theta||^2 + \frac{d+||\theta||^2}{(||\theta||^2-t^2)^2}(\sqrt{d} + ||\theta||^2)^2 \\ &\asymp d + ||\theta||^2 + \frac{d+||\theta||^2}{||\theta||^4}||\theta||^4 \\ &\asymp d + ||\theta||^2. \end{aligned}$$

此情形的证明至此完成。

情形 3：考虑情形 $0 < ||\theta|| < 2t$。那么

$$\begin{aligned} E(||Z||^2 - \alpha_t(d) \,|\, ||Z||^2 \geq d + t^2) &= E(||g||^2 - d \,|\, ||g+\theta||^2 \geq d + t^2) + ||\theta||^2 - t^2 \\ &\leq E(||g||^2 - d \,|\, ||g+\theta||^2 \geq d + t^2) + 3t^2 \\ &\leq E(||g||^2 - d \,|\, ||g||^2 \geq d + t^2) + 3t^2 \\ &= \alpha_t(d) - d + 3t^2 \\ &\lesssim \beta\sqrt{d} + t^2 \\ &\asymp t^2 \end{aligned}$$

其中我们使用了引理 21。因此，将上述界与式（56）结合并代入式（55），得到

$$\mathrm{Var}(Y) \lesssim d + ||\theta||^2 + t^4 \asymp t^4.$$

证明完毕。 □

引理 15. *Let L^* be the universal positive constant from Lemma 21. There exists universal positive constants D and C^* such that if $d \geq D$, then for any $1 \leq \beta \leq L^*\sqrt{d}$ and $x > 0$, we have*

$$P\left\{\sum_{j=1}^{n}\left(||Z_j||^2 - \alpha_t(d)\right)_{\mathbb{I}\{||Z_j||^2 \geq d+t^2\}} \geq C^*\left(\sqrt{xnt^4 e^{-\frac{ct^4}{d}}} + \frac{d}{t^2}x\right)\right\} \leq e^{-x}$$

where $Z_1, \ldots\ldots, Z_n \overset{iid}{\sim} N(0, I_d)$ $t^2 = \beta\sqrt{d}$, and $\alpha_t(d)$ is given by (13). Here, $c > 0$ is a universalconstant.

Proof. 引理11 的证明只需稍作修改：调用引理 12 替代引理 8，并对 $\lambda < \frac{\beta}{2(\beta C^* + \sqrt{d})}$ 取下确界。 □

B χ^2_d 分布的性质

定理 7 (Bernstein 不等式 - 定理 2.8。1 [46])。 *Let $Y_1,\ldots,Y_d$ be independent mean-zero subexponential random variables. Then, for every $u \geq 0$, we have*

$$P\left\{\left|\sum_{i=1}^{d} Y_i\right| \geq u\right\} \leq 2\exp\left(-c\min\left(\frac{u^2}{\sum_{i=1}^{d}||X_i||^2_{\psi_1}}, \frac{u}{\max_i ||X_i||_{\psi_1}}\right)\right)$$

where $c > 0$ is a universal constant and ψ_2, ψ_1 denote the subgaussian and subexponential norms respectively (see (2.13) and (2.21) of [46])

推论 1. *Suppose $Z \sim N(0, I_d)$. If $u \geq 0$, then*

$$P\left\{||Z||^2 \geq d + u\right\} \leq 2\exp\left(-c\min\left(\frac{u^2}{d}, u\right)\right)$$

where $c > 0$ is a universal constant.

引理 16 (引理 1 [31]). *Let $Z_1,\ldots,Z_d \overset{iid}{\sim} N(0,1)$. If $\lambda_1 \geq \ldots \geq \lambda_d > 0$ and $x > 0$, then*

$$P\left\{\sum_{j=1}^{d}\lambda_j Z_j^2 \geq \sum_{j=1}^{d}\lambda_j + 2\sqrt{x\sum_{j=1}^{d}\lambda_j^2} + 2\lambda_1 x\right\} \leq e^{-x}.$$

引理 17 (推论 3 [51]). *Suppose $Z \sim N(0, I_d)$. There exist universal constants $C, c > 0$ suchthat*

$$P\left\{||Z||^2 \geq d + u\right\} \geq C\exp\left(-c\min\left(\frac{u^2}{d}, u\right)\right)$$

for all $u > 0$.

引理 18（[28]）。 *Let f_d and F_d respectively denote the probability density and cumulative distribution functions of the χ^2_d distribution. Then the following relations hold,*

(i) $tf_d(t) = df_{d+2}(t)$,

(ii) $t^2 f_d(t) = d(d+2) f_{d+4}(t)$,

(iii) $f'_{d+2}(t) = \frac{f_d(t) - f_{d+2}(t)}{2}$,

(iv) $(1 - F_{d+2}(t)) - (1 - F_d(t)) = \frac{e^{-t/2} t^{d/2}}{2^{d/2}\Gamma(d/2+1)}$.

引理 19。 *Let f_d and F_d respectively denote the probability density and cumulative distribution functions of the χ^2_d distribution. Suppose $d > 2$. If $t \geq d$, then $2f_d(t) \leq 1 - F_d(t)$。*

Proof. 根据中值定理，对于任意 $r \geq t$,

$$\begin{aligned}\inf_{x \geq d} \frac{\frac{\partial}{\partial x}(1 - F_d(x))}{2f'_d(x)} &\leq \frac{(1 - F_d(t)) - (1 - F_d(r))}{2f_d(t) - 2f_d(r)} \\ &= \frac{\frac{1-F_d(t)}{2f_d(t)} - \frac{1-F_d(r)}{2f_d(t)}}{1 - \frac{f_d(r)}{f_d(t)}}.\end{aligned}$$

注意到 $\lim_{r\to\infty} 1 - F_d(r) = \lim_{r\to\infty} f_d(r) = 0$。因此取 $r \to \infty$ 可得

$$\inf_{x \geq d} \frac{\frac{\partial}{\partial x}(1 - F_d(x))}{2f'_d(x)} \leq \frac{1 - F_d(t)}{2f_d(t)}.$$

现在我们来求左侧的下确界。对于 $x \geq d$，注意到应用引理18 可得

$$\begin{aligned}\frac{\frac{\partial}{\partial x}(1 - F_d(x))}{2f'_d(x)} &= -\frac{f_d(x)}{f_{d-2}(x) - f_d(x)} \\ &= -\frac{1}{\frac{f_{d-2}(x)}{f_d(x)} - 1} \\ &= -\frac{1}{\frac{xf_{d-2}(x)}{xf_d(x)} - 1} \\ &= -\frac{1}{\frac{d-2}{x} - 1} \\ &= \frac{1}{1 - \frac{d-2}{x}}.\end{aligned}$$

由于 $x \geq d > d - 2$，可知

$$\inf_{x \geq d} \frac{\frac{\partial}{\partial x}(1 - F_d(x))}{2f'_d(x)} = 1.$$

因此，我们可以立即得出 $2f_d(t) \leq 1 - F_d(t)$，如所需。 □

L引理 20. *Let F_d denote the cumulative distribution function of the χ^2_d distribution. If $x \geq 0$, then*

$$1 - F_d(x) = \frac{1}{\Gamma\left(\frac{d}{2}\right)} \int_{x/2}^{\infty} t^{d/2-1} e^{-t}\, dt = Q\left(\frac{d}{2}, \frac{x}{2}\right)$$

where Q is the upper incomplete gamma function defined in Theorem 8.

Proof. 在对概率密度函数积分时，通过变量替换即可直接得到该结果。

□

引理 21. *Suppose d is larger than a sufficiently large universal constant. There exist universal positive constants L^* and C^* such that the following holds. If $1 \leq \beta \leq L^*\sqrt{d}$ and $t^2 = \beta\sqrt{d}$, then*

$$\alpha_t(d) \leq d + C^*\beta\sqrt{d}$$

where $\alpha_t(d)$ is given by (13).

Proof. 为简化记号，令 $r^2 = d + t^2$。令 f_d 和 F_d 分别表示 χ^2_d 分布的概率密度和累积分布函数。我们将在证明后文中选取通用常数 L^*。根据引理18，有

$$\begin{aligned}
\alpha_t(d) &= \frac{\int_{r^2}^{\infty} z f_d(z)\,dz}{1 - F_d(r^2)} \\
&= \frac{\int_{r^2}^{\infty} df_{d+2}(z)\,dz}{1 - F_d(r^2)} \\
&= d\left(1 + \frac{\left(1 - F_{d+2}(r^2)\right) - \left(1 - F_d(r^2)\right)}{1 - F_d(r^2)}\right) \\
&= d\left(1 + \frac{r^d e^{-r^2/2}}{2^{d/2}\Gamma\left(\frac{d}{2}+1\right)} \cdot \frac{1}{1 - F_d(r^2)}\right) \\
&= d\left(1 + \left(\frac{r^2}{2}\right)^{d/2} \frac{e^{-r^2/2}}{\Gamma\left(\frac{d}{2}+1\right)} \cdot \frac{1}{1 - F_d(r^2)}\right)
\end{aligned}$$

重新整理各项并应用斯特林近似（即 $\Gamma(x+1) \sim \sqrt{2\pi x}\left(\frac{x}{e}\right)^x \text{as} x \to \infty$），得到

$$\left(\frac{r^2}{2}\right)^{d/2} \frac{e^{-r^2/2}}{\Gamma\left(\frac{d}{2}+1\right)} \le \frac{1+c}{\sqrt{\pi d}} \exp\left(\frac{d}{2}\log\left(\frac{r^2}{2}\right) - \frac{r^2}{2} - \frac{d}{2}\log\left(\frac{d}{2}\right) + \frac{d}{2}\right)$$

其中通用常数 $c > 0$ 因为 d 大于一个充分大的通用常数。应用引理 20 和推论 2 可得

$$1 - F_d(r^2) = Q\left(\frac{d}{2}, \frac{r^2}{2}\right) \ge \left(1 - \Phi(\eta\sqrt{a})\right) - \frac{e^{-a\eta^2/2}}{\sqrt{2\pi}} \cdot \frac{c^{**}}{\sqrt{d}}$$

其中 c^{**} 是一个通用常数。注意，由于 d 大于一个充分大的通用常数，因此 $\eta\sqrt{a}$ 也大于一个充分大的通用常数。利用 $1 - \Phi(x) = \frac{1}{x\sqrt{2\pi}} e^{-x^2/2}(1 + o(1))$ 当 $x \to \infty$ 时的事实，可得

$$1 - F_d(r^2) \ge \frac{1}{\sqrt{2\pi}} e^{-\frac{a\eta^2}{2}} \left(\frac{c^*}{\eta\sqrt{a}} - \frac{c^{**}}{\sqrt{d}}\right)$$

对于某个普适正常数 c^*。考虑

$$\frac{\eta^2}{2} = \mu - \log(1+\mu) = \frac{\frac{r^2}{2}}{\frac{d}{2}} - 1 - \log\left(\frac{r^2}{2}\right) + \log\left(\frac{d}{2}\right).$$

因此，

$$-\frac{\eta^2 a}{2} = -\left(\frac{r^2}{2} - \frac{d}{2} - \frac{d}{2}\log\left(\frac{r^2}{2}\right) + \frac{d}{2}\log\left(\frac{d}{2}\right)\right) = -\frac{r^2}{2} + \frac{d}{2} + \frac{d}{2}\log\left(\frac{r^2}{2}\right) - \frac{d}{2}\log\left(\frac{d}{2}\right).$$

因此，我们得到如下界

$$1 - F_d(r^2) \ge \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{r^2}{2} + \frac{d}{2} + \frac{d}{2}\log\left(\frac{r^2}{2}\right) - \frac{d}{2}\log\left(\frac{d}{2}\right)\right) \cdot \left(\frac{c^*}{\eta\sqrt{a}} - \frac{c^{**}}{\sqrt{d}}\right).$$

进一步考虑不等式 $x - \log(1+x) \le \frac{x^2}{2} \le x^2$，可得

$$\eta\sqrt{a} = \sqrt{\frac{d}{2}} \cdot \sqrt{2\left(\frac{\beta}{\sqrt{d}} - \log\left(1 + \frac{\beta}{\sqrt{d}}\right)\right)} = \sqrt{d\left(\frac{\beta}{\sqrt{d}} - \log\left(1 + \frac{\beta}{\sqrt{d}}\right)\right)} \le \beta.$$

因此，我们有

$$1 - F_d(r^2) \ge \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{r^2}{2} + \frac{d}{2} + \frac{d}{2}\log\left(\frac{r^2}{2}\right) - \frac{d}{2}\log\left(\frac{d}{2}\right)\right)\left(\frac{c^*}{\beta} - \frac{c^{**}}{\sqrt{d}}\right).$$

令 $L^* := \frac{c^*}{2c^{**}}$。作此选择后，我们有 $\beta \le \sqrt{d} \cdot \frac{c^*}{2c^{**}}$，因此 $\left(\frac{c^*}{\beta} - \frac{c^{**}}{\sqrt{d}}\right) \ge \frac{c^*}{2\beta}$。所以

re,

$$\left(\frac{r^2}{2}\right)^{d/2} \frac{e^{-r^2/2}}{\Gamma\left(\frac{d}{2}+1\right)} \cdot \frac{1}{1-F_d(r^2)} \le \frac{1+c}{\sqrt{\pi d}} \cdot \frac{2\beta\sqrt{2\pi}}{c^*} \cdot \frac{\exp\left(\frac{d}{2}\log\left(\frac{r^2}{2}\right) - \frac{r^2}{2} - \frac{d}{2}\log\left(\frac{d}{2}\right) + \frac{d}{2}\right)}{\exp\left(-\frac{r^2}{2} + \frac{d}{2} + \frac{d}{2}\log\left(\frac{r^2}{2}\right) - \frac{d}{2}\log\left(\frac{d}{2}\right)\right)} \le C^* \frac{\beta}{\sqrt{d}}$$

对于某个普适正常数 C^*，因此我们有

$$\alpha_t(d) \le d\left(1 + \frac{C^*\beta}{\sqrt{d}}\right) = d + C^*\beta\sqrt{d}$$

正如所需。 □

定理 8 (不完全伽马函数的均匀展开 [43]). *For $a > 0$ and $x \ge 0$ realnumbers, define the upper incomplete gamma function*

$$Q(a,x) := \frac{1}{\Gamma(a)} \int_x^\infty t^{a-1} e^{-t}\, dt.$$

Further define $\lambda = \frac{x}{a}, \mu = \lambda - 1$, and $\eta = \sqrt{2(\mu - \log(1+\mu))}$. Then Q admits an asymptotic series expansion in a which is uniform in $\eta \in \mathbb{R}$. In other words, for any integer $N \ge 0$, we have

$$Q(a,x) = \left(1 - \Phi\left(\eta\sqrt{a}\right)\right) + \frac{e^{-\frac{a\eta^2}{2}}}{\sqrt{2\pi a}} \sum_{k=0}^{N} c_k(\eta) a^{-k} + \mathrm{Rem}_N(a,\eta)$$

where the remainder term satisfies

$$\lim_{a\to\infty} \sup_{\eta\in\mathbb{R}} \left| \frac{\mathrm{Rem}_N(a,\eta)}{\frac{e^{-\frac{a\eta^2}{2}}}{\sqrt{2\pi a}} c_N(\eta) a^{-N}} \right| = 0.$$

Here, Φ denotes the cumulative distribution function of the standard normal distribution.

定理 9 (定理 1 见 [42]). *Consider the setting of Theorem 8. The coefficient c_0 is given by* $c_0(\eta) = \frac{1}{\mu} - \frac{1}{\eta}$.

推论 2。 *Consider the setting of Theorem 8. For any $a > 0$ and $x \ge 0$, we have*

$$Q(a,x) = \left(1 - \Phi\left(\eta\sqrt{a}\right)\right) + \frac{e^{-\frac{a\eta^2}{2}}}{\sqrt{2\pi a}} \left(\frac{1}{\mu} - \frac{1}{\eta}\right) + \mathrm{Rem}_0(a,\eta)$$

where

$$\lim_{a\to\infty} \sup_{\eta\in\mathbb{R}} \left| \frac{\mathrm{Rem}_0(a,\eta)}{\frac{e^{-\frac{a\eta^2}{2}}}{\sqrt{2\pi a}} \left(\frac{1}{\mu} - \frac{1}{\eta}\right)} \right| = 0.$$

Consequently, if a is larger than some universal positive constant and $\mu > 0$, we have

$$\left|Q(a,x) - \left(1 - \Phi\left(\eta\sqrt{a}\right)\right)\right| \le \frac{Ce^{-\frac{a\eta^2}{2}}}{\sqrt{2\pi a}}$$

where C is some universal positive constant.

Proof. 前两个显示式直接由定理8 和9得出。为证明最后一个显示式，我们必须证明当 $\left|\frac{1}{\mu} - \frac{1}{\eta}\right| \lesssim 1$ 时 $\mu > 0$成立。首先，考虑泰勒展开式

$$\log(1+\mu) = \mu - \frac{\mu^2}{2(1+\xi)^2}$$

其中， ξ 是介于 0 和 μ之间的某个点。因此，

$$\sqrt{2(\mu - \log(1+\mu))} = \frac{\mu}{1+\xi}.$$

Thus

$$\left|\frac{1}{\mu}-\frac{1}{\eta}\right|=\left|\frac{1}{\mu}-\frac{1+\xi}{\mu}\right|=\frac{\xi}{\mu}\leq 1$$

由于 ξ 介于 0 和 μ之间。因此，只需令 a 大于某个通用常数，并令 C 为足够大的通用常数，即可得到推论陈述中的最终显示式。

□

引理22。 *Let* $Z\sim N(0,I_d)$. *If* $t\geq 0$, *then*

$$E(||Z||^2\,|\,||Z||^2\geq d+t^2)\leq d+C^*(t^2\vee d)$$

where C^* *is a positive universal constant.*

Proof. 我们将在证明的最后选取一个通用常数 $L\geq 1$，因此目前暂将其留作未定值。令 f_d 表示 χ^2_d 分布的概率密度函数。注意到

$$\begin{aligned}E(||Z||^2\,|\,||Z||^2\geq d+t^2)&=E(||Z||^2\mathbb{I}_{\{||Z||^2\leq d+Lt^2\}}\,|\,||Z||^2\geq d+t^2)+E(||Z||^2\mathbb{I}_{\{||Z||^2>d+Lt^2\}}\,|\,||Z||^2\geq d+t^2)\\&\leq d+Lt^2+\sqrt{E(||Z||^4\,|\,||Z||^2\geq d+t^2)}\sqrt{P\{||Z||^2>d+Lt^2\}}.\end{aligned}$$

根据推论 1，存在一个通用常数 c_1 使得

$$\sqrt{P\{||Z||^2>d+Lt^2\}}\leq\sqrt{2}\exp\left(-Lc_1\min\left(\frac{t^4}{d},t^2\right)\right).$$

这里我们使用了 $L\geq 1$。进一步考虑

$$\begin{aligned}E(||Z||^4\,|\,||Z||^2\geq d+t^2)&=\frac{\int_{d+t^2}^{\infty}z^2f_d(z)\,dz}{P\{||Z||^2\geq d+t^2\}}\\&\leq\frac{E(||Z||^4)}{P\{||Z||^2\geq d+t^2\}}\\&=\frac{d(d+2)}{P\{||Z||^2\geq d+t^2\}}.\end{aligned}$$

因此，根据引理 17 我们有

$$E(||Z||^2\,|\,||Z||^2\geq d+t^2)\leq d+Lt^2+dC\cdot\frac{\exp\left(-Lc_1\min\left(\frac{t^4}{d},t^2\right)\right)}{\exp\left(-c_2\min\left(\frac{t^4}{d},t^2\right)\right)}$$

其中 C 和 c_2 为普适正常数。取 $L:=\frac{c_2}{c_1}\vee 1$ 即完成证明。 □

引理23。 *Let* $Z\sim N(0,I_d)$. *If* $t\geq 0$, *then*

$$E(||Z||^4\,|\,||Z||^2\geq d+t^2)\leq d^2+C^*(t^4\vee d^2)$$

where C^* *is a positive universal constant.*

Proof. 我们将在证明的最后选取一个通用常数 $L\geq 1$，因此目前将其留作未定。令 f_d 表示 χ^2_d 分布的概率密度函数。考虑

$$\begin{aligned}&E(||Z||^4\,|\,||Z||^2\geq d+t^2)\\&=E(||Z||^4\mathbb{I}_{\{||Z||^2\leq d+Lt^2\}}\,|\,||Z||^2\geq d+t^2)+E(||Z||^4\mathbb{I}_{\{||Z||^2\leq d+Lt^2\}}\,|\,||Z||^2\geq d+t^2)\\&\leq d^2+2Lt^2d+L^2t^4+\sqrt{E(||Z||^8\,|\,||Z||^2\geq d+t^2)}\sqrt{P\{||Z||^2>d+Lt^2\}}.\end{aligned}$$

根据推论 1，存在一个普适正常数 c_1 使得

$$\sqrt{P\{||Z||^2>d+Lt^2\}}\leq\sqrt{2}\exp\left(-Lc_1\min\left(\frac{t^4}{d},t^2\right)\right).$$

这里我们使用了 $L \geq 1$。进一步考虑

$$\begin{aligned} E(||Z||^8 \,|\, ||Z||^2 \geq d + t^2) &= \frac{\int_{d+t^2}^{\infty} z^4 f_d(z)\, dz}{P\{||Z||^2 \geq d + t^2\}} \\ &\leq \frac{E(||Z||^8)}{P\{||Z||^2 \geq d + t^2\}} \\ &= \frac{d(d+2)(d+4)(d+6)}{P\{||Z||^2 \geq d + t^2\}}. \end{aligned}$$

因此，由引理 17 可得

$$E(||Z||^4 \,|\, ||Z||^2 \geq d + t^2) \leq d^2 + 2Lt^2 d + L^2 t^4 + d^2 C \frac{\exp\left(-Lc_1 \min\left(\frac{t^4}{d}, t^2\right)\right)}{\exp\left(-c_2 \min\left(\frac{t^4}{d}, t^2\right)\right)}$$

其中 $C > 0$ 是一个通用常数。取 $L > \frac{c_2}{c_1} \vee 1$ 即可完成证明。 □

引理 24. *Let $Z \sim N(0, I_d)$. If $t \geq 0$, then*

$$\mathrm{Var}(||Z||^2 \,|\, ||Z||^2 \geq d + t^2) \lesssim t^4 + d e^{-C \min\left(\frac{t^4}{d}, t^2\right)}$$

for some universal positive constant C.

Proof. 在证明中，我们将使用一个通用常数 $L \geq 1$；其取值将在最后给出。令 f_d 表示 χ^2_d 分布的概率密度。注意到

$$\begin{aligned} &\mathrm{Var}(||Z||^2 \,|\, ||Z||^2 \geq d + t^2) \\ &= \mathrm{Var}(||Z||^2 - d \,|\, ||Z||^2 \geq d + t^2) \\ &\leq E((||Z||^2 - d)^2 \,|\, ||Z||^2 \geq d + t^2) \\ &= E((||Z||^2 - d)^2 \mathbb{I}_{\{||Z||^2 \leq d + Lt^2\}} \,|\, ||Z||^2 \geq d + t^2) + E((||Z||^2 - d)^2 \mathbb{I}_{\{||Z||^2 > d + Lt^2\}} \,|\, ||Z||^2 \geq d + t^2) \\ &\leq L^2 t^4 + \sqrt{E((||Z||^2 - d)^4 \,|\, ||Z||^2 \geq d + t^2)} \sqrt{P\{||Z||^2 \geq d + Lt^2\}}. \end{aligned}$$

由推论 1，存在一个普适正常数 c_1 使得

$$\sqrt{P\{|Z||^2 \geq d + Lt^2\}} \leq \sqrt{2} \exp\left(-c_1 L \min\left(\frac{t^4}{d}, t^2\right)\right).$$

这里，我们用到了 $L \geq 1$。控制住这一项后，再注意到

$$\begin{aligned} E((||Z||^2 - d)^4 \,|\, ||Z||^2 \geq d + t^2) &= \frac{\int_{d+t^2}^{\infty} (z-d)^2 f_d(z)\, dz}{P\{||Z||^2 \geq d + t^2\}} \\ &\leq \frac{E((||Z||^2 - d)^4)}{P\{||Z||^2 \geq d + t^2\}} \\ &= \frac{12d(d+4)}{P\{||Z||^2 \geq d + t^2\}}. \end{aligned}$$

因此，由引理17 可得

$$\sqrt{E((||Z||^2 - d)^4 \,|\, ||Z||^2 \geq d + t^4)} \leq \frac{C^* d}{\exp\left(-c_2 \min\left(\frac{t^4}{d}, t^2\right)\right)}.$$

取 $L > \frac{c_2}{c_1} \vee 1$，我们有

$$\begin{aligned} \mathrm{Var}(||Z||^2 \,|\, ||Z||^2 \geq d + t^2) &\leq L^2 t^4 + \frac{C^* d}{\exp\left(-c_2 \min\left(\frac{t^4}{d}, t^2\right)\right)} \cdot \sqrt{2} \exp\left(-c_1 L \min\left(\frac{t^4}{d}, t^2\right)\right) \\ &\leq L^2 t^4 + d \cdot C^* \sqrt{2} \exp\left(-(c_1 L - c_2) \min\left(\frac{t^4}{d}, t^2\right)\right). \end{aligned}$$

这正是所需的结果。 □

C 杂项

C.1 极小极大

引理 25。 *We have*

$$\Gamma_{\mathcal{H}} = \mu_{\nu_{\mathcal{H}}} \vee \frac{\sqrt{(\nu_{\mathcal{H}} - 1)\log\left(1 + \frac{p}{s^2}\right)}}{n}.$$

Consequently, $\Gamma_{\mathcal{H}} \le \mu_{\nu_{\mathcal{H}}-1}$.

Proof. 考虑

$$\begin{aligned}
\Gamma_{\mathcal{H}} &= \max_{\nu<\nu_{\mathcal{H}}} \left\{ \mu_\nu \wedge \frac{\sqrt{\nu \log\left(1 + \frac{p}{s^2}\right)}}{n} \right\} \vee \max_{\nu \ge \nu_{\mathcal{H}}} \left\{ \mu_\nu \wedge \frac{\sqrt{\nu \log\left(1 + \frac{p}{s^2}\right)}}{n} \right\} \\
&= \max_{\nu<\nu_{\mathcal{H}}} \left\{ \frac{\sqrt{\nu \log\left(1 + \frac{p}{s^2}\right)}}{n} \right\} \vee \max_{\nu \ge \nu_{\mathcal{H}}} \left\{ \mu_\nu \wedge \frac{\sqrt{\nu \log\left(1 + \frac{p}{s^2}\right)}}{n} \right\} \\
&= \frac{\sqrt{(\nu_{\mathcal{H}} - 1)\log\left(1 + \frac{p}{s^2}\right)}}{n} \vee \mu_{\nu_{\mathcal{H}}}.
\end{aligned}$$

这里，我们利用了特征值的排序以及 $\mu_{\nu_{\mathcal{H}}} \le \frac{\sqrt{\nu_{\mathcal{H}} \log\left(1+\frac{p}{s^2}\right)}}{n}$ 这一事实，得到了第二项。~~根据 $\nu_{\mathcal{H}}$ 的定义，界 $\Gamma_{\mathcal{H}} \le \mu_{\nu_{\mathcal{H}}-1}$ 可由 $\mu_{\nu_{\mathcal{H}}} \le \mu_{\nu_{\mathcal{H}}-1}$ 和 $\frac{\sqrt{(\nu_{\mathcal{H}}-1)\log\left(1+\frac{p}{s^2}\right)}}{n} \le \mu_{\nu_{\mathcal{H}}-1}$ 立即得到。~~ □

Proof of Lemma 1. 该结果是引理 2的一个特例。 □

C.2 自适应

引理 26. *Suppose* $a > 0$. *We have*

$$\Gamma_{\mathcal{H}}(s,a) = \mu_{\nu_{\mathcal{H}}(s,a)} \vee \frac{\sqrt{(\nu_{\mathcal{H}}(s,a) - 1)\log\left(1 + \frac{pa}{s^2}\right)}}{n}.$$

Consequently, $\Gamma_{\mathcal{H}}(s,a) \le \mu_{\nu_{\mathcal{H}}(s,a)-1}$.

Proof. 注意到

$$\begin{aligned}
\Gamma_{\mathcal{H}}(s,a) &= \max_{\nu<\nu_{\mathcal{H}}(s,a)} \left\{ \mu_\nu \wedge \frac{\sqrt{\nu \log\left(1 + \frac{pa}{s^2}\right)}}{n} \right\} \vee \max_{\nu \ge \nu_{\mathcal{H}}(s,a)} \left\{ \mu_\nu \wedge \frac{\sqrt{\nu \log\left(1 + \frac{pa}{s^2}\right)}}{n} \right\} \\
&= \max_{\nu<\nu_{\mathcal{H}}(s,a)} \left\{ \frac{\sqrt{\nu \log\left(1 + \frac{pa}{s^2}\right)}}{n} \right\} \vee \max_{\nu \ge \nu_{\mathcal{H}}(s,a)} \left\{ \mu_\nu \wedge \frac{\sqrt{\nu \log\left(1 + \frac{pa}{s^2}\right)}}{n} \right\} \\
&= \frac{\sqrt{(\nu_{\mathcal{H}}(s,a) - 1)\log\left(1 + \frac{pa}{s^2}\right)}}{n} \vee \mu_{\nu_{\mathcal{H}}(s,a)}.
\end{aligned}$$

这里，我们利用特征值的排序以及 $\mu_{\nu_{\mathcal{H}}(s,a)} \le \frac{\sqrt{\nu_{\mathcal{H}}(s,a) \log\left(1+\frac{pa}{s^2}\right)}}{n}$ 得到第二项。~~根据定义，要对 $\Gamma_{\mathcal{H}} \le \mu_{\nu_{\mathcal{H}}(s,a)-1}$ 进行界定，可直接由 $\mu_{\nu_{\mathcal{H}}} \le \mu_{\nu_{\mathcal{H}}(s,a)-1}$ 和 $\frac{\sqrt{(\nu_{\mathcal{H}}(s,a)-1)\log\left(1+\frac{pa}{s^2}\right)}}{n} \le \mu_{\nu_{\mathcal{H}}(s,a)-1}$ 得出 $\nu_{\mathcal{H}}(s,a)$。~~ □

Proof of Lemma 2. 我们首先证明第二个不等式。由于 $\log\left(1 + \frac{pa}{s^2}\right) \le \frac{n}{2}$ 和 $\mu_1 = 1$，因此

立即可得 $\nu_{\mathcal{H}}(s,a) \geq 2$。因此 e,

$$\begin{aligned}\frac{\sqrt{\nu_{\mathcal{H}}(s,a)\log\left(1+\frac{pa}{s^2}\right)}}{n} &\leq \sqrt{2}\frac{\sqrt{(\nu_{\mathcal{H}}(s,a)-1)\log\left(1+\frac{pa}{s^2}\right)}}{n} \\ &\leq \sqrt{2}\left(\mu_{\nu_{\mathcal{H}}(s,a)} \vee \frac{\sqrt{(\nu_{\mathcal{H}}(s,a)-1)\log\left(1+\frac{pa}{s^2}\right)}}{n}\right) \\ &\leq \sqrt{2}\Gamma_{\mathcal{H}}(s,a)\end{aligned}$$

如所需。

现在证明第一个不等式。对于任意 $\nu \geq \nu_{\mathcal{H}}(s,a)$，根据特征值的排序，我们有

$$\mu_\nu \wedge \frac{\sqrt{\nu\log\left(1+\frac{pa}{s^2}\right)}}{n} \leq \mu_\nu \leq \mu_{\nu_{\mathcal{H}}}(s,a) \leq \frac{\sqrt{\nu_{\mathcal{H}}(s,a)\log\left(1+\frac{pa}{s^2}\right)}}{n}.$$

对于任意 $\nu < \nu_{\mathcal{H}}(s,a)$，根据 $\nu_{\mathcal{H}}(s,a)$ 的定义，可得

$$\mu_\nu \wedge \frac{\sqrt{\nu\log\left(1+\frac{pa}{s^2}\right)}}{n} = \frac{\sqrt{\nu\log\left(1+\frac{pa}{s^2}\right)}}{n} \leq \frac{\sqrt{\nu_{\mathcal{H}}(s,a)\log\left(1+\frac{pa}{s^2}\right)}}{n}.$$

由于对所有 ν，我们已经证明

$$\mu_\nu \wedge \frac{\sqrt{\nu\log\left(1+\frac{pa}{s^2}\right)}}{n} \leq \frac{\sqrt{\nu_{\mathcal{H}}(s,a)\log\left(1+\frac{pa}{s^2}\right)}}{n},$$

对 ν 取最大值，得到

$$\Gamma_{\mathcal{H}}(s,a) \leq \frac{\sqrt{\nu_{\mathcal{H}}(s,a)\log\left(1+\frac{pa}{s^2}\right)}}{n}$$

如所需。 □

Proof of Lemma 3. 为简化记号，记作 $a = \mathscr{A}_{\mathcal{H}}$。根据假设 $\log(1+pa) \leq \frac{n}{2}$ 和 $\mu_1 = 1$，我们有 $\nu_{\mathcal{H}}(s,a) \geq 2$，对所有 s均成立。根据 $\nu_{\mathcal{H}}(s,a)$ 的定义，我们有

$$\mu_{\nu_{\mathcal{H}}(s,a)} \leq \frac{\sqrt{\nu_{\mathcal{H}}(s,a)\log\left(1+\frac{pa}{s^2}\right)}}{n}$$

and

$$\mu_{\nu_{\mathcal{H}}(s,a)-1} > \frac{\sqrt{(\nu_{\mathcal{H}}(s,a)-1)\log\left(1+\frac{pa}{s^2}\right)}}{n}$$

定义

$$\delta_s := \inf\left\{\delta > 0 : \mu_{\nu(s,a)-1} \leq \frac{\sqrt{(\nu_{\mathcal{H}}(s,a)-1)\log\left(1+\frac{p(a+\delta)}{s^2}\right)}}{n}\right\}.$$

由于前倒数第二个公式中的不等式是严格的，因此显然 $\delta_s > 0$ 对每个 s 都成立。此外，根据 $\nu_{\mathcal{H}}$ 的定义，可知 $\nu_{\mathcal{H}}(s,a+\delta) > \nu_{\mathcal{H}}(s,a) - 1$ 对所有 $\delta < \delta_s$ 均成立。又由于 $\nu_{\mathcal{H}}$ 关于其第二个自变量递减，因此可得 $\nu_{\mathcal{H}}(s,a+\delta) = \nu_{\mathcal{H}}(s,a)$，对所有 $\delta < \delta_s$ 均成立。令$\delta^* := 1 \wedge \min_{s\in[p]} \delta_s$，立即可得 $\mathscr{V}_a = \mathscr{V}_{a+\delta}$ 对所有 $\delta < \delta^*$ 均成立。根据 $\mathscr{A}_{\mathcal{H}}$ 的定义，对于任意 $\delta < \delta^*$，可得

$$\mathscr{A}_{\mathcal{H}} = a \leq \log(e|\mathscr{V}_a|) = \log(e|\mathscr{V}_{a+\delta}|) < a + \delta \leq 2\mathscr{A}_{\mathcal{H}}$$

其中，我们使用了 $\mathscr{A}_{\mathcal{H}} \geq 1$ 和 $\delta \leq 1$ 得到了最后的不等式。证明完毕。 □

C.3 技术

命题 5 (Ingster-Suslina 方法 [24])。 *Suppose* $\Sigma \in \mathbb{R}^{d\times d}$ *is a positive definite matrix and* $\Theta \subset \mathbb{R}^d$ *is a parameter space. If* π *is a probability distribution supported on* Θ, *then*

$$\chi^2\left(\int N(\theta,\Sigma)\,\pi(d\theta)\middle\|\, N(0,\Sigma)\right) = E\left(\exp\left(\left\langle\theta,\Sigma^{-1}\widetilde{\theta}\right\rangle\right)\right) - 1$$

where $\theta,\widetilde{\theta}\overset{iid}{\sim}\pi$ *and* $\chi^2(\cdot||\cdot)$ *denotes the* χ^2*-divergence defined in Section 1.1.*

引理 27 ([45]). *If* P, Q *are probability measures on a measurable space* $(\mathcal{X},\mathcal{A})$ *with* $P \ll Q$, *then* $d_{TV}(P,Q) \le \frac{1}{2}\sqrt{\chi^2(P||Q)}$.

引理 28. *If* Y *is distributed according to the hypergeometric distribution with probability mass function* $P\{Y=k\} = \frac{\binom{s}{k}\binom{n-s}{t-k}}{\binom{n}{t}}$ *for* $0 \le k \le s\wedge t$, *then* $E(\exp(\lambda Y)) \le \left(1-\frac{s}{n}+\frac{s}{n}e^{\lambda}\right)^t$ *for* $\lambda > 0$.

推论 3 ([10]). *If* Y *is distributed according to the hypergeometric distribution with probability mass function* $P\{Y=k\} = \frac{\binom{s}{k}\binom{n-s}{s-k}}{\binom{n}{s}}$ *for* $0 \le k \le s \le n$, *then* $E(Y) = \frac{s^2}{n}$ *and* $E(\exp(\lambda Y)) \le \left(1-\frac{s}{n}+\frac{s}{n}e^{\lambda}\right)^s$ *for* $\lambda > 0$.