

Minimax Signal Detection in Sparse Additive Models

Subhodh Kotekal and Chao Gao

University of Chicago

Abstract

Sparse additive models are an attractive choice in circumstances calling for modelling flexibility in the face of high dimensionality. We study the signal detection problem and establish the minimax separation rate for the detection of a sparse additive signal. Our result is nonasymptotic and applicable to the general case where the univariate component functions belong to a generic reproducing kernel Hilbert space. Unlike the estimation theory, the minimax separation rate reveals a nontrivial interaction between sparsity and the choice of function space. We also investigate adaptation to sparsity and establish an adaptive testing rate for a generic function space; adaptation is possible in some spaces while others impose an unavoidable cost. Finally, adaptation to both sparsity and smoothness is studied in the setting of Sobolev space, and we correct some existing claims in the literature.

1 Introduction

In the interest of interpretability, computation, and circumventing the statistical curse of dimensionality plaguing high dimensional regression, structure is often assumed on the true regression function. Indeed, it might plausibly be argued that sparse linear regression is the distinguishing export of modern statistics. Despite its popularity, circumstances may call for more flexibility to capture nonlinear effects of the covariates. Striking a balance between flexibility and structure, Hastie and Tibshirani [20] proposed generalized additive models (GAMs) as a natural extension to the vaunted linear model. In a GAM, the regression function admits an additive decomposition of univariate (nonlinear) component functions. However, as in the linear model, the sample size must outpace the dimension for consistent estimation. Following modern statistical instinct, a sparse additive model is compelling [29, 33, 35, 38, 39, 50]. The regression function admits an additive decomposition of univariate functions for which only a small subset are nonzero; it is the combination of a GAM and sparsity.

To fix notation, consider the p -dimensional Gaussian white noise model

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

for $x \in [0, 1]^p$. Though it may be more faithful to practical data analysis to consider the nonparametric regression model $Y_i = f(X_i) + \epsilon_i$ (e.g. as in [38]), the Gaussian white noise model is convenient as it avoids distracting technicalities while maintaining focus on the statistical essence. Indeed, the nonparametric statistics literature has a long history of studying the white noise model to understand theoretical limits, relying on well-known asymptotic equivalences [5, 40] which imply, under various conditions, that mathematical results obtained in one model can be ported over to the other model. As our focus is theoretical rather than practical, we follow in this tradition. The generalized additive model asserts the regression function is of the form $f(x) = \sum_{j=1}^p f_j(x_j)$ with $f_1, \dots, f_p$ being univariate functions belonging to some function space $\mathcal{H}$. Likewise, the sparse additive model asserts

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

for some unknown support set S of size s denoting the active covariates.

Most of the existing literature has addressed estimation of sparse additive models, primarily in the nonparametric regression setting and with $\mathcal{H}$ being a reproducing kernel Hilbert space. After a series of works [29, 33, 35, 39], Raskutti et al. [38] (see also [29]) established that a penalized M -estimator achieves the minimax estimation rate under various choices of the reproducing kernel Hilbert space $\mathcal{H}$. Yuan and Zhou [50] establish minimax estimation rates under a notion of approximate sparsity. As is now seen as typical of estimation theory, the powerful framework of empirical processes is brought down to bear on their proofs. Some articles have also addressed generalizations of the sparse additive model. For example, the authors of [49] consider, among other structures, an additive signal $f(x) = \sum_{j=1}^k f_j(x)$ where each component function f_j is actually a multivariate function depending on at most s_j many coordinates and is α_j -Hölder. The authors go on to derive minimax rates that handle heterogeneity in the smoothness indices and the sparsities of the coordinate functions; as a taste of their results, they show, in a particular regime and under some conditions, the rate $kn^{-\frac{2\alpha}{2\alpha+s}} + ks \log(\frac{p}{s})/n$ in the special, homogeneous case where $s_j = s$ and $\alpha_j = \alpha$ for all j . Recently, the results of [3] show certain deep neural networks can achieve the minimax estimation rate for sparse k -way interaction models. The k -way interaction model is also known as nonparametric ANOVA. To give an example, the sparse 2-way interaction model assumes $f(x) = \sum_{j \in S_1} f_j(x_j) + \sum_{(k,l) \in S_2} f_{kl}(x_k, x_l)$ where the sets of active variables S_1 and interactions S_2 have small cardinalities. When the f_j are β_1 -Hölder and the f_{kl} are β_2 -Hölder, [3] establishes, under some conditions and up to factors logarithmic in n , the rate $s_1(n^{-\frac{2\beta_1}{2\beta_1+1}} + (\log p)/n) + s_2(n^{-\frac{2\beta_2}{2\beta_2+2}} + (\log p)/n)$.

The literature has much less to say on the problem of signal detection

$$H_0 : f \equiv 0, \quad (2)$$

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

where $\mathcal{F}_s$ is the class of sparse additive signals given by (4). Adopting a minimax perspective [6, 21–24], the goal is to determine the smallest rate ε as a function of the sparsity level s , the dimension p , the sample size n , and the function space $\mathcal{H}$ such that consistent detection of the alternative against the null is possible.

Though to a much lesser extent than the corresponding estimation theory, optimal testing rates have been established in various high dimensional settings other than sparse additive models. The most canonical setup, the Gaussian sequence model, is perhaps the most studied [2, 7, 8, 10, 12, 14, 18, 24, 30, 34, 41]. Optimal rates have also been established in linear regression [1, 26, 37] and other models [13, 36]. A common motif is that optimal testing rates exhibit different phenomena from optimal estimation rates.

Returning to (2)–(3), the only directly relevant works in the literature are Ingster and Lepski’s article [25] and a later article by Gayraud and Ingster [17]. Ingster and Lepski [25] consider a sparse multichannel model which, after a transformation to sequence space, is closely related to (2)–(3). Adopting an asymptotic setting and exclusively choosing $\mathcal{H}$ to be a Sobolev space, they establish asymptotic minimax separation rates. However, their results only address the regimes $p = O(s^2)$ and $s = O(p^{1/2-\delta})$ for a constant $\delta \in (0, 1/2)$. Their result does not precisely pin down the rate near the phase transition $s \asymp \sqrt{p}$. In essence, their testing procedure in the sparse regime is to apply a Bonferroni correction to a collection of χ^2 -type tests, one test per each of the p coordinates. Thus, a gap in their rate near $\sqrt{p}$ is unsurprising. In the dense regime, a single χ^2 -type test is used, as is typical in minimax testing literature. Ingster and Lepski [25] also address adaptation to sparsity as well as adaptation both the sparsity and the smoothness. Ingster and Gayraud [17] consider sparse additive models rather than a sparse multichannel model but make the same choice of $\mathcal{H}$ and work in an asymptotic setup. They establish sharp constants for the sparse case $s = p^{1/2-\delta}$ via a Higher Criticism type testing statistic. Throughout the paper, we make comparisons of our results to primarily [25] as it was the first article to establish rates. Our results do not address the question of sharp constants.

Our paper's main contributions are the following. First, adopting a nonasymptotic minimax testing framework as initiated in [2], we establish the nonasymptotic minimax separation rate for (2)-(3) for any configuration of the sparsity s , dimension p , sample size n , and a generic function space $\mathcal{H}$. Notably, we do not restrict ourselves to Sobolev (or Besov) space as in [17, 25]. The test procedure we analyze involves thresholded χ^2 statistics, following a strategy employed in other sparse signal detection problems [8–10, 30, 34].

Our second contribution is to establish an adaptive testing rate for a generic function space. Typically, the sparsity level is unknown, and it is of practical interest to have a methodology which can accommodate a generic $\mathcal{H}$. Interestingly, some choices of $\mathcal{H}$ do not involve any cost of adaptation, that is, the minimax rate can be achieved without knowing the sparsity. Our rate's incurred adaptation cost turns out to be a delicate function of $\mathcal{H}$, thus extending Ingster and Lepski's focus on Sobolev spaces [25]. Even in the Sobolev case, our result extends upon their article; near the regime $s \asymp \sqrt{p}$, our test provides finer detail by incurring a cost involving only $\log \log \log p$ instead of $\log \log p$ as incurred by their test. In principle, our result can be used to reverse the process and find a space $\mathcal{H}$ for which this adaptive rate incurs a given adaptation cost.

Finally, adaptation to both sparsity and smoothness is studied in the context of Sobolev space. We identify an issue with and correct a claim made by [25].

1.1 Notation

The following notation will be used throughout the paper. For $p \in \mathbb{N}$, let $[p] := \{1, \dots, p\}$. For $a, b \in \mathbb{R}$, denote $a \vee b := \max\{a, b\}$ and $a \wedge b = \min\{a, b\}$. Denote $a \lesssim b$ to mean there exists a universal constant $C > 0$ such that $a \leq Cb$. The expression $a \gtrsim b$ means $b \lesssim a$. Further, $a \asymp b$ means $a \lesssim b$ and $b \lesssim a$. The symbol $\langle \cdot, \cdot \rangle$ denotes the usual inner product in Euclidean space and $\langle \cdot, \cdot \rangle_F$ denotes the Frobenius inner product. The total variation distance between two probability measures P and Q on a measurable space $(\mathcal{X}, \mathcal{A})$ is defined as $d_{TV}(P, Q) := \sup_{A \in \mathcal{A}} |P(A) - Q(A)|$. If Q is absolutely continuous with respect to P , the χ^2 -divergence is defined as $\chi^2(Q \| P) := \int_{\mathcal{X}} \left(\frac{dQ}{dP} - 1 \right)^2 dP$. For a finite set S , the notation $|S|$ denotes the cardinality of S . Throughout, iterated logarithms will be used (e.g. expressions like $\log \log \log p$ and $\log \log(np)$). Without explicitly stating so, we will take such an expression to be equal to some universal constant if otherwise it would be less than one. For example, $\log \log \log p$ should be understood to be equal to a universal constant when $p < e^{e^e}$.

1.2 Setup

1.2.1 Reproducing Kernel Hilbert Space (RKHS)

Following the literature (e.g. [47]), $\mathcal{H}$ will denote a reproducing kernel Hilbert space (RKHS). Before discussing main results, basic properties of RKHSs are reviewed [47]. Suppose $\mathcal{H} \subset L^2([0, 1])$ is a reproducing kernel Hilbert space (RKHS) with associated inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$. There exists a symmetric function $K : [0, 1] \times [0, 1] \rightarrow \mathbb{R}_+$ called a *kernel* such that for any $x \in [0, 1]$ we have (1) the function $K(\cdot, x) \in \mathcal{H}$ and (2) for all $f \in \mathcal{H}$ we have $f(x) = \langle f, K(\cdot, x) \rangle_{\mathcal{H}}$. Mercer's theorem (Theorem 12.20 in [47]) guarantees that the kernel K admits an expansion in terms of eigenfunctions $\{\psi_k\}_{k=1}^{\infty}$, namely $K(x, x') = \sum_{k=1}^{\infty} \mu_k \psi_k(x) \psi_k(x')$. To give examples, the kernel $K(x, x') = 1 + \min\{x, x'\}$ defines the first-order Sobolev space with eigenvalue decay $\mu_k \asymp k^{-2}$, and the kernel $K(x, x') = \exp\left(-\frac{(x-x')^2}{2}\right)$ exhibits eigenvalue decay $\mu_k \asymp e^{-ck \log k}$ (see [47] for a more detailed review).

Without loss of generality, we order the eigenvalues $\mu_1 \geq \mu_2 \geq \dots \geq 0$. The eigenfunctions $\{\psi_k\}_{k=1}^{\infty}$ are orthonormal in $L^2([0, 1])$ under the usual L^2 inner product $\langle \cdot, \cdot \rangle_{L^2}$ and the inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ satisfies $\langle f, g \rangle_{\mathcal{H}} = \sum_{k=1}^{\infty} \mu_k^{-1} \langle f, \psi_k \rangle_{L^2} \langle g, \psi_k \rangle_{L^2}$ for $f, g \in \mathcal{H}$.

1.2.2 Parameter space

The parameter space which will be used throughout the paper is defined in this section. Suppose $\mathcal{H}$ is an RKHS. Recall we are interested in sparse additive signals $f(x) = \sum_{j \in S} f_j(x_j)$ for some

sparsity pattern $S \subset [p]$. Following [17, 38], for convenience we assume $\int_0^1 f_j(t) dt = 0$ for all j . This assumption is mild and can be relaxed; further discussion can be found in Section 5.1. Letting $\mathbf{1} \in L^2([0, 1])$ denote the constant function equal to one, consider that $\mathcal{H}_0 := \mathcal{H} \cap \text{span}\{\mathbf{1}\}^\perp$ is a closed subspace of $\mathcal{H}$. Hence, $\mathcal{H}_0$ is also an RKHS. We will put aside $\mathcal{H}$ (along with its eigenfunctions and eigenvalues) and only work with $\mathcal{H}_0$. Let $\{\psi_k\}_{k=1}^\infty$ and $\{\mu_k\}_{k=1}^\infty$ denote its associated eigenfunctions and eigenvalues respectively. Following [2], we assume $\mu_1 = 1$ also for convenience; this can easily be relaxed.

For each subset $S \subset [p]$, define

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

The condition on the RKHS norm enforces regularity. Define the parameter space

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

for each $1 \leq s \leq p$.

Following [17, 25], it is convenient to transform (1) from function space to sequence space. The tensor product $L^2([0, 1])^{\otimes p}$ admits the orthonormal basis $\{\Phi_\ell\}_{\ell \in \mathbb{N}^p}$ with $\Phi_\ell(t) = \prod_{j=1}^p \psi_{\ell_j}(t_j)$ for $t \in [0, 1]^p$. For ease of notation, denote $\phi_{j,k}(t) = \psi_k(t_j)$ for $k \in \mathbb{N}$ and $j \in [p]$. Define the random variables

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

where $\theta_{k,j} = \int_0^1 \psi_k(x) f_j(x) dx$. The assumption $\int_0^1 f_j(t) dt = 0$ for all j is used here. Note by orthogonality that $\{X_{k,j}\}_{k \in \mathbb{N}, j \in [p]}$ is a collection of independent random variables. The notation $\Theta \in \mathbb{R}^{\mathbb{N} \times p}$ will frequently be used to denote the full matrix of coefficients. The notation $\Theta_j \in \mathbb{R}^{\mathbb{N}}$ will also be used to denote the j th column of Θ . For $f \in \mathcal{F}_s$, the corresponding set of coefficients is

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

Note the parameter spaces $\mathcal{F}_s$ and $\mathcal{T}_s$ are in correspondence, and we will frequently write f and Θ freely in the same context without comment. The understanding is f is a function and Θ is its corresponding basis coefficients. The notation E_f, P_f, E_Θ , and P_Θ will be used to denote expectations and probability measures with respect to the denoted parameters.

1.2.3 Problem

As described earlier, given an observation from (1) the signal detection problem (2)-(3) is of interest. The goal is to characterize the nonasymptotic minimax separation rate $\varepsilon_{\mathcal{H}}^* = \varepsilon_{\mathcal{H}}^*(p, s, n)$.

Definition 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 $0 < c < c_\eta$,

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

The nonasymptotic minimax testing framework was initiated by Baraud [2] and characterizes (up to universal factors) the fundamental statistical limit of the detection problem. The framework is nonasymptotic in the sense that the conditions for the minimax separation rate hold for any configuration of the problem parameters.

Since $\|f\|_2^2 = \sum_{j=1}^p \sum_{k=1}^\infty \theta_{k,j}^2$, the problem can be equivalently formulated in sequence space as

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

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

This testing problem, with the parameter space (6), is interesting its own right outside the sparse additive model and RKHS context. Indeed, the testing problem (7)-(8) is essentially a group-sparse extension of the detection problem in ellipses considered in Baraud's seminal article [2]. In fact, this interpretation was actually our initial motivation to study the detection problem. The connection to sparse additive models was a later consideration; similar to the way in which the later article [17] considers sparse additive models when building upon the earlier, fundamental work [25] dealing with a sparse multichannel (essentially group-sparse) model. Taking the perspective of a sequence problem has a long history in nonparametric regression [2, 5, 15, 24, 25, 27, 40, 48] due to not only its fundamental connections but also its advantage in distilling the problem to its essence and dispelling technical distractions. Our results can be exclusively read (and readers are encouraged to do so) in the context of the sequence problem (7)-(8).

2 Minimax rates

We begin by describing some high-level and natural intuition before informally stating our main result in Section 2.1. Section 2.2 contains the development of some key quantities. In Sections 2.3 and 2.4, we formally state minimax lower and upper bounds respectively. In Section 2.5, some special cases illustrating the general result are discussed.

2.1 A naive ansatz

A first instinct is to look to previous results for context in an attempt to make a conjecture regarding the optimal testing rate. To illustrate how this line of thinking might proceed, consider the classical Gaussian white noise model on the unit interval in one dimension,

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

for $x \in [0, 1]$. Assume f lives inside the unit ball of a reproducing kernel Hilbert space $\mathcal{H}$ and thus admits a decomposition in the associated orthonormal basis with a coefficient vector $\theta \in \ell^2(\mathbb{N})$ living in an ellipsoid determined by the kernel's eigenvalues $\mu_1 \geq \mu_2 \geq \dots \geq 0$. The optimal rate of estimation is given by Birgé and Massart [4]

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

Baraud [2] established that the minimax separation rate for the signal detection problem

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

is given by

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

In both estimation and testing, the maximizer ν^* can be conceptualized as the correct truncation order. Specifically for testing, Baraud's procedure [2], working in sequence space, rejects H_0 when $\sum_{k=1}^{\nu^*} X_k^2 - \frac{\nu^*}{n} \gtrsim \frac{\sqrt{\nu^*}}{n}$. The data for $k > \nu^*$ are not used, and it is in this sense ν^* is understood as a truncation level. To illustrate these results, the rates for Sobolev spaces with smoothness α are $\epsilon_{\text{est}}^2 \asymp n^{-\frac{2\alpha}{2\alpha+1}}$ and $\epsilon_{\text{test}}^2 \asymp n^{-\frac{4\alpha}{4\alpha+1}}$ since $\mu_\nu \asymp \nu^{-2\alpha}$. By now, these nonasymptotic results of [2, 4] are well known and canonical.

Moving to the setting of sparse additive signals in the model (1), Raskutti et al. [38] derive the nonasymptotic minimax rate of estimation

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

While their upper bound holds for any choice of $\mathcal{H}$ (satisfying some mild conditions) and sparsity level s , they only obtain a matching lower bound when $s = o(p)$ and when the unit ball of $\mathcal{H}$ has logarithmic or polynomial scaling metric entropy. This rate obtained by Raskutti et al. [38] is pleasing. It is quite intuitive to see the term $\frac{s \log p}{n}$ due to the sparsity in the parameter space. The term $s \epsilon_{\text{est}}^2$ is the natural rate for estimating s many univariate functions in $\mathcal{H}$ as if sparsity pattern were known. Notably, there is no interaction between the choice of $\mathcal{H}$ and the sparsity structure. The sparsity term $\frac{s \log p}{n}$ is independent of $\mathcal{H}$ and the estimation term $s \epsilon_{\text{est}}^2$ is dimension free.

One might intuit that this lack of interaction is a general phenomenon. Instinct may suggest that signal detection in sparse additive models is also just s many instances of a univariate nonparametric detection problem plus the problem of a detecting a signal which is nonzero on an unknown sparsity pattern. Framing it as two distinct problems, one might conjecture the optimal testing rate should be $s \epsilon_{\text{test}}^2$ plus the s -sparse detection rate. Collier et al. [10] provide a natural candidate for the s -sparse testing rate, namely the rate $\frac{s \log(\frac{p}{s^2})}{n}$ for $s < \sqrt{p}$ and $\frac{\sqrt{p}}{n}$ for $s \geq \sqrt{p}$.

However, this is not the case as a quick glance at [25] falsifies the conjecture for the case of Sobolev $\mathcal{H}$. Though quick to dispel hopeful thinking, [25] expresses little about the interaction between sparsity and $\mathcal{H}$. Our result explicitly captures this nontrivial interaction for a generic $\mathcal{H}$. We show the minimax separation rate is given by

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

The rate bears some resemblance to the sparse testing rate of Collier et al. [10] and the nonparametric testing rate of Baraud [2], but the combination of the two is not *a priori* straightforward. At this point in discussion, not enough has been developed to explain how the form of the rate arises. Various features of the rate will be commented on later on in the paper.

2.2 Preliminaries

In this section, some key pieces are defined.

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

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

Note, since $\mu_1 = 1$, it follows $\Gamma_{\mathcal{H}} \gtrsim \frac{1}{n}$. It is readily seen from (9) there are two broad regimes to consider. When $n \lesssim \log(1 + p/s^2)$, we have $\varepsilon_{\mathcal{H}}^*(p, s, n)^2 \asymp s$. In the regime $n \gtrsim \log(1 + p/s^2)$, the rate is more complicated. The first regime is really a trivial regime since any signal $f \in \mathcal{F}_s$ must satisfy $\|f\|_2^2 \leq s$ by virtue of $\mu_1 = 1$. Therefore, the degenerate test which always accepts H_0 vacuously detects sparse additive signals with $\|f\|_2^2 \geq 2s$. Hence, the upper bound $\varepsilon_{\mathcal{H}}^*(p, s, n)^2 \lesssim s$ is trivially achieved. It turns out a matching lower bound can be proved which establishes that the regime $n \lesssim \log(1 + p/s^2)$ is fundamentally trivial; see Section 2.3 for a formal statement.

More generally, the form (9) is useful when discussing lower bounds. A different form is more convenient when discussing upper bounds, and it is a form which is familiar in the context of [10, 11].

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

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

As the next lemma shows, $\nu_{\mathcal{H}}$ is essentially the solution to the maximization problem in the definition of $\Gamma_{\mathcal{H}}$. Drawing an analogy to the result of Baraud [2] described in Section 2.1, $\nu_{\mathcal{H}}$ can be conceptualized as the correct order of truncation accounting for the dimension and sparsity.

Lemma 1. If $\log \left(1 + \frac{p}{s^2}\right) \leq \frac{n}{2}$, then $\Gamma_{\mathcal{H}} \leq \frac{\sqrt{\nu_{\mathcal{H}} \log \left(1 + \frac{p}{s^2}\right)}}{n} \leq \sqrt{2} \Gamma_{\mathcal{H}}$.

With Lemma 1, the testing rate can be expressed as

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

The condition $\log \left(1 + \frac{p}{s^2}\right) \lesssim n$ in Lemma 1 is natural in light of the triviality which occurs when $n \lesssim \log \left(1 + \frac{p}{s^2}\right)$.

2.3 Lower bound

In this section, a formal statement of a lower bound on the minimax separation rate is given. Define

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

First, it is shown that the testing problem is trivial if $\log \left(\frac{p}{s^2}\right) \gtrsim n$.

Proposition 1 (Triviality). Suppose $1 \leq s \leq p$. Suppose $\kappa > 0$ and $\log \left(1 + \frac{p}{s^2}\right) \geq \kappa n$. If $\eta \in (0, 1)$, then for any $0 < c < 1 \wedge \sqrt{\kappa} \wedge \sqrt{\kappa \log(1 + 4\eta^2)}$ we have

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

Proposition 1 asserts that in order to achieve small testing error, a necessary condition is that $\|f\|_2 \geq C\sqrt{s}$ for a sufficiently large $C > 0$. To see why Proposition 1 is a statement about triviality, observe that $\mathcal{F}_s \subset \{f \in L^2([0, 1]^p) : \|f\|_2 \leq \sqrt{s}\}$. To reiterate plainly, all potential $f \in \mathcal{F}_s$ in the model (1) live inside the ball of radius $\sqrt{s}$. There are essentially no functions $f \in \mathcal{F}_s$ that have detectable norm when $\log \left(\frac{p}{s^2}\right) \gtrsim n$, and so the problem is essentially trivial.

The lower bound construction is straightforward. Working in sequence space, consider an alternative hypothesis with a prior π in which a draw $\Theta \sim \pi$ is obtained by drawing a size s subset $S \subset [p]$ uniformly at random and setting $\theta_{k,j} = \rho$ if $k = 1$ and $j \in S$ or setting $\theta_{k,j} = 0$ otherwise. The value of ρ determines the separation between the null and alternative hypotheses since $\|\Theta\|_F^2 = s\rho^2$. However, ρ must respect some constraints to ensure π is supported on the parameter space and that it is impossible to distinguish the null and alternative hypotheses with vanishing error. Observe this construction places us in a one-dimensional sparse sequence model (since $\theta_{k,j} = 0$ for $k \geq 2$), which is precisely the setting of [10]. From their results, it is seen we must have $\rho^2 \lesssim \log(1 + \frac{p}{s^2})/n$. To ensure π is supported on the parameter space, we must have $\sum_{k=1}^{\infty} \theta_{k,j}^2/\mu_k \leq 1$ for all $j \in [p]$. Since $\mu_1 = 1$, it follows $\sum_{k=1}^{\infty} \theta_{k,j}^2/\mu_k = \theta_{1,j}^2 \leq \rho^2$, and so the constraint $\rho \lesssim 1$ must be enforced. When $\log(1 + p/s^2) \gtrsim n$, only the second condition $\rho \lesssim 1$ is binding, and so the largest separation we can achieve is $\|\Theta\|_F^2 \asymp s$. Hence, the problem is trivial in this regime.

To ensure non-triviality, it will be assumed $\log(1 + \frac{p}{s^2}) \leq \frac{n}{2}$. The choice of the factor $1/2$ is only for convenience and is not essential. In fact, the condition $\log(1 + \frac{p}{s^2}) < n$ would always suffice for our purposes, and the condition $\log(1 + \frac{p}{s^2}) \leq n$ would also suffice for $n > 1$. The following theorem establishes $\varepsilon_{\mathcal{H}}^*(p, s, n)^2 \gtrsim \psi(p, s, n)^2$.

Theorem 1. *Suppose $1 \leq s \leq p$ and $\log(1 + \frac{p}{s^2}) \leq \frac{n}{2}$. 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).

The lower bound is proved via Le Cam's two point method (also known as the method of "fuzzy" hypotheses [45]) which is standard in the minimax hypothesis testing literature [24]. The prior distribution employed in the argument is a natural construction. Namely, the active set (i.e. the size s set of nonzero coordinate functions) is selected uniformly at random and the corresponding nonzero coordinate functions are drawn from the usual univariate nonparametric prior employed in the literature [2, 24].

2.4 Upper bound

In this section, testing procedures are constructed and formal statements establishing rate-optimality are made. The form of the rate in (10) is a more convenient target, and should be kept in mind when reading the statements of our upper bounds.

2.4.1 Hard thresholding in the sparse regime

In this section, the sparse regime $s < \sqrt{p}$ is discussed. For any $d \in \mathbb{N}$ and $j \in [p]$, define $E_j(d) = n \sum_{k \leq d} X_{k,j}^2$ where the data $\{X_{k,j}\}_{k \in \mathbb{N}, j \in [p]}$ is defined via transformation to sequence space (5). For any $r \geq 0$, define

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

where

$$\alpha_r(d) := \frac{E(\|g\|^2 \mathbb{1}_{\{\|g\|^2 \geq d+r^2\}})}{P\{\|g\|^2 \geq d+r^2\}} \quad (13)$$

where $g \sim N(0, I_d)$. Note $\alpha_r(d)$ is a conditional expectation under the null hypothesis. The random variable $T_r(d)$ will be used as a test statistic. Such a statistic was first defined by Collier et al. [10], and similar statistics have been successfully employed in other signal detection problems [8, 9, 30, 34]. However, all previous works in the literature have only used this statistic with $d = 1$. For our problem, it is necessary to take growing d . Consequently, a more refined analysis is necessary, and essentially new phenomena appear in the upper bound as we discuss later. As noted in Section 2.2,

the quantity $\nu_{\mathcal{H}}$ can be conceptualized of as the correct truncation order, that is, it turns out the correct choice is $d \asymp \nu_{\mathcal{H}}$. The choice of r is more complicated, and in fact there are two separate regimes depending on the size of d . The regime in which $d \gtrsim \log(1 + \frac{p}{s^2})$ is referred to as the “bulk” regime. The complementary regime $d \lesssim \log(1 + \frac{p}{s^2})$ is referred to as the “tail” regime.

Proposition 2 (Bulk). *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 $1 \leq s < \sqrt{p}$ and $\log(1 + \frac{p}{s^2}) \leq K_3^2 d$. Set*

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

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

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

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

Proposition 3 (Tail). *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 $1 \leq s < \sqrt{p}$ and $\log(1 + \frac{p}{s^2}) \geq K_3^2 d$. Set*

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

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

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

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

Propositions 3 and 2 thus imply that, for $s < \sqrt{p}$, the minimax separation rate is upper bounded by $s \log(1 + \frac{p}{s^2})/n + s \sqrt{\nu_{\mathcal{H}} \log(1 + \frac{p}{s^2})}/n$. By Lemma 1, it follows that

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

under the condition that $\log(1 + \frac{p}{s^2}) \leq \frac{n}{2}$. As established by Proposition 1 in Section 2.3, a condition like this is essential to avoid triviality.

Propositions 3 and 2 reveal an interesting phase transition in the minimax separation rate at the point

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

This phase transition phenomena is driven by the tail behavior of χ_d^2 . Consider that under the null distribution $E_j(d) \sim \chi_d^2$ for all j , and so the statistic $T_r(d)$ is the sum of p independent and thresholded χ_d^2 random variables. By a well-known lemma of Laurent and Massart [31] (also from Bernstein’s inequality up to constants), for any $u > 0$ we have

$$P \left\{ \chi_d^2 - d \geq 2\sqrt{du} + 2u \right\} \leq e^{-u}. \quad (14)$$

Roughly speaking, $\chi_d^2 - d$ exhibits subgaussian-type deviation in the “bulk” $u \lesssim \sqrt{d}$ and subexponential-type deviation in the “tail” $u \gtrsim \sqrt{d}$. Consequently, $T_r(d)$ should be intuited as a sum of thresholded subgaussian random variables when $r \lesssim \sqrt{d}$, and as a sum of thresholded subexponential random variables when $r \gtrsim \sqrt{d}$.

Examining (14), in the “tail” $u \gtrsim \sqrt{d}$ it follows $2\sqrt{du} + 2u \asymp u$, which no longer exhibits dependence on d . Analogously, in Proposition 3 the threshold is taken as $r \gtrsim \sqrt{d}$ and so the resulting rate exhibits no dependence on d and consequently no dependence on $\mathcal{H}$. On the other hand, in the

“bulk” $u \lesssim \sqrt{d}$ it follows $2\sqrt{du} + 2u \asymp \sqrt{du}$. Analogously, in Proposition 2 the threshold is taken as $r \lesssim \sqrt{d}$ and so the resulting rate indeed depends on d and thus on $\mathcal{H}$.

2.4.2 χ^2 tests in the dense regime

The situation is less complicated in the dense regime $s \geq \sqrt{p}$ as it suffices to use the χ^2 testing statistic $T := n \sum_{j=1}^p \sum_{k \leq \nu_{\mathcal{H}}} X_{k,j}^2$.

Proposition 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

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

As in the sparse case, (10) Lemma 1 asserts $\tau(p, s, n)^2 \asymp s\Gamma_{\mathcal{H}}$ provided that the condition $\log(1 + \frac{p}{s^2}) \leq n/2$ is satisfied.

2.5 Special cases

Having formally stated lower and upper bounds for the minimax separation rate, it is informative to explore a number of special cases and witness a variety of phenomena. Throughout the illustrations it will be assumed $\log(1 + \frac{p}{s^2}) \leq n/2$ as discussed earlier.

2.5.1 Sobolev

Taking the case $\mu_k \asymp k^{-2\alpha}$ as emblematic of Sobolev space with smoothness $\alpha > 0$, we obtain the minimax separation rate

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

It is useful to compare to the rates obtained by Ingster and Lepski (Theorem 2 in [25]) (see also [17]) although their choice of parameter is space is slightly different from that considered in this paper. In the sparse case $s = O(p^{1/2-\delta})$ for a constant $\delta \in (0, 1/2]$, their rate is

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

In the dense regime $p = O(s^2)$, their rate (Theorem 1 in [25]) is $s \left(\frac{ns}{\sqrt{p}} \right)^{-\frac{4\alpha}{4\alpha+1}}$. Quick algebra verifies that our rate indeed matches Ingster and Lepski’s rate in these sparsity regimes.

In the sparse regime, the strange looking phase transition at $\log p \asymp n^{\frac{1}{2\alpha+1}}$ in their rate now has a coherent explanation in view of our result. The situation $\log p \lesssim n^{\frac{1}{2\alpha+1}}$ corresponds to the “bulk” regime in which case $T_r(d)$ from (12) behaves like a sum of thresholded subgaussian random variables. On the other side where $\log p \gtrsim n^{\frac{1}{2\alpha+1}}$, subexponential behavior is exhibited by $T_r(d)$. In fact, our result gives more detail beyond $s \lesssim p^{1/2-\delta}$. Assume only $s < \sqrt{p}$ (for example, $s = \frac{\sqrt{p}}{\log p}$ is allowed now). Then the phase transition between the “bulk” and “tail” regimes actually occurs at $\log(1 + \frac{p}{s^2}) \asymp n^{\frac{1}{2\alpha+1}}$.

2.5.2 Finite dimension

Consider a finite dimensional situation, that is $1 = \mu_1 = \mu_2 = \dots = \mu_m > \mu_{m+1} = \mu_{m+2} = \dots = 0$ for some positive integer m . Function spaces exhibiting this kind of structure include linear functions, finite polynomials, and generally RKHSs based on kernels with finite rank. If $m < \frac{n^2}{\log(1 + \frac{p}{s^2})}$, the minimax separation rate is

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

In the sparse regime, the phase transition between the bulk and tail regimes occurs at $\log(1 + \frac{p}{s^2}) \asymp m$.

2.5.3 Exponential decay

As another example, consider exponential decay of the eigenvalues $\mu_k = c_1 e^{-c_2 k^\gamma}$ where c_1, c_2 are universal constants and $\gamma > 0$. Such decay is a feature of RKHSs based on Gaussian kernels. The minimax separation rate is

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

In the sparse regime, the phase transition between the bulk and tail regimes occurs at $\log^\gamma(1 + \frac{p}{s^2}) \asymp \log n$. The minimax separation rate is quite close to the finite dimensional rate, which is sensible as RKHSs based on Gaussian kernels are known to be fairly “small” nonparametric function spaces [47].

3 Adaptation

Thus far the sparsity parameter s has been assumed known, and the tests constructed make use of this information. In practice, the statistician is typically ignorant of the sparsity level and so it is of interest to understand whether adaptive tests can be furnished. In this section, we will establish an adaptive testing rate which accommodates a generic $\mathcal{H}$, and it turns out to exhibit a cost for adaptation which depends delicately on the function space.

To the best of our knowledge, Spokoiny’s article was the first to demonstrate an unavoidable cost for adaptation in a signal detection problem [41]. Later work established unavoidable costs for adaptive testing across a variety of situations (see [24] and references therein). This early work largely focused on adapting to an unknown smoothness parameter in a univariate nonparametric regression setting. More recently adaptation to sparsity in high dimensional models has been studied. In many problems, one can adapt to sparsity without cost in the rate (nor in the constant for some problems) [1, 14, 18, 26, 30, 34]. In the context of sparse additive models in Sobolev space, Ingster and Lepski [25] (see also [17]) consider adaptation, and we discuss their results in Section 3.4.

3.1 Preliminaries

To formally state our result, some slight generalizations of the concepts found in Section 2.2 are needed.

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

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

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

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

Note $\nu_{\mathcal{H}}(s, a)$ is decreasing in a and $\Gamma_{\mathcal{H}}(s, a)$ is increasing in a . As discussed in Section 2.2, two different forms are useful when discussing the separation rate, and Lemma 1 facilitated that discussion. The following lemma is a slight generalization and has the same purpose.

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

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

At a high level, the central issue with adaptation lies in the selection of an estimator's or a test's hyperparameter (such as the bandwidth or penalty). Typically, there is some optimal choice but it requires knowledge of an unknown parameter (e.g. smoothness or sparsity) in order to pick it. In the current problem, the optimal choice of ν is unknown since the sparsity level s is unknown.

In adaptive testing, the typical strategy is to fix a grid of different values, construct a test for each potential value of the hyperparameter, and for detection use the test which takes the maximum over this collection of tests. Typically, a geometric grid is chosen, and the logarithm of the grid's cardinality reflects a cost paid by the testing procedure [25, 34, 41]. It turns out this high level intuition holds for signal detection in sparse additive models, but the details of how to select the grid are not direct since a generic $\mathcal{H}$ must be accommodated. For $a > 0$, define

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

Define

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

and define

$$\mathcal{V}_{\mathcal{H}} := \mathcal{V}_{\mathcal{A}_{\mathcal{H}}}. \quad (16)$$

The grid $\mathcal{V}_{\mathcal{H}}$ will be used. It is readily seen that $\mathcal{A}_{\mathcal{H}}$ is finite as the crude bound $\mathcal{A}_{\mathcal{H}} \leq \log(ep)$ is immediate. The following lemma shows that $\mathcal{A}_{\mathcal{H}}$ is, in essence, a fixed point.

Lemma 3. If $\log \left(1 + p\mathcal{A}_{\mathcal{H}}\right) \leq \frac{n}{2}$, then $\mathcal{A}_{\mathcal{H}} \leq \log(e|\mathcal{V}_{\mathcal{H}}|) \leq 2\mathcal{A}_{\mathcal{H}}$.

It turns out an adaptive testing procedure can be constructed which pays a price determined by $\mathcal{A}_{\mathcal{H}}$ (equivalently $\log(e|\mathcal{V}_{\mathcal{H}}|)$ up to constant factors). To elaborate, assume $\log \left(1 + p\mathcal{A}_{\mathcal{H}}\right) \leq n/2$. We will construct an adaptive test which achieves

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

The form of this separation rate is very similar to the minimax rate, except that the phase transition has shifted to $s \asymp \sqrt{p\mathcal{A}_{\mathcal{H}}}$ and a cost involving $\mathcal{A}_{\mathcal{H}}$ is incurred. The value of $\mathcal{A}_{\mathcal{H}}$ can vary quite a bit with the choice of $\mathcal{H}$, and a few examples are illustrated in Section 3.4.

Our choice of the geometric grid yielding the fixed point characterization of Lemma 3 may not be immediately intuitive. It can be understood at a high level by noticing there are two competing factors at play. First, fix $a \geq 1$ and note it can be conceptualized to represent a target cost from scanning over some grid. To achieve that target cost, we should use the truncation levels $\nu_{\mathcal{H}}(s, a)$ for each sparsity s . Now, consider that we will scan over the geometric grid $\mathcal{V}_a$ obtained from the truncation levels. If that geometric grid has cardinality which would incur a cost less than a , then it would appear strange that we are scanning over a smaller grid to achieve a target cost associated to a larger grid. It is intuitive that we might try to aim for a lower target cost $a' < a$ instead. But this

changes our grid to $\mathcal{V}_{a'}$ which now has a different cardinality, and so the same logic applies to a' . A similar line of reasoning applies if the grid had happened to be too large. Therefore, we intuitively want to select a such that $\mathcal{V}_a$ has the proper cardinality to push the cost a upon us. Hence, we are seeking a kind of fixed point in this sense.

3.2 Lower bound

The formulation of a lower bound is more delicate than one might initially anticipate. To illustrate the subtlety, suppose we have a candidate adaptive rate $s \mapsto \epsilon(s)$. It is tempting to consider the following lower bound formulation; for any $\eta \in (0, 1)$, there exists $c_\eta > 0$ such that for any $0 < c < c_\eta$,

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

While extremely natural, this criterion has an issue. In particular, if there exists $\bar{s}$ such that $\epsilon(\bar{s}) \asymp \epsilon^*(\bar{s})$ where $s \mapsto \epsilon^*(s)$ denotes the minimax rate, then the above criterion is trivially satisfied. One simply lower bounds the maximum over all s with the choice $s = \bar{s}$ and then invokes the minimax lower bound. To exaggerate, the candidate rate $\epsilon(1) = \epsilon^*(1)$ and $\epsilon(s) = \infty$ for $s \geq 2$ satisfies the above criterion. Furthermore, from an upper bound perspective there is a trivial test which achieves this candidate rate. The absurdity from this lower bound criterion would then force us to conclude this candidate rate gives an adaptive testing rate.

To avoid such absurdities, we use a different lower bound criterion. The key issue in the formulation is that, in the domain of maximization for the Type II error, one must not include any s for which the candidate rate is of the same order as the minimax rate.

Definition 6. Suppose $s \mapsto \varepsilon_{\text{adapt}}(p, s, n)$ 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 $s \mapsto \varepsilon_{\text{adapt}}(p, s, n)$ 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 $0 < c < c_\eta$,

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

for all p and n .

(ii) Either

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

or

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

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

Intuitively, this criterion is nontrivial only when there are no sparsity levels in the chosen reference sets for which the candidate rate is the same order as the minimax rate. More explicitly, there is no sequence $\{s_{p,n}\}_{p,n \in \mathbb{N}}$ with $s_{p,n} \in \mathcal{S}_{p,n}$ along which the candidate and minimax rates match (up to constants). This criterion avoids trivialities such as the one described earlier. In Definition 6, formalizing the notion of “order” requires speaking in terms of sequences, though it may appear unfamiliar and clunky at first glance. Definition 6 is not new, but rather a direct port to the testing context of the notion of an *adaptive rate of convergence on the scale of classes* from [44] in the estimation setting (see also [11] for an application of this notion to linear functional estimation in the sparse Gaussian sequence model).

The condition (18) simply requires the candidate rate to match the minimax rate if (17) does not

hold. In this way, the definition allows for the possibility of the minimax rate to be an adaptive rate (in those cases where no cost for adaptation needs to be paid).

We now state a lower bound with respect to this criterion. Define the candidate rate

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

Here, $\mathcal{A}_{\mathcal{H}}$ is defined in (15). Examining this candidate, it is possible one may run into absurdities for $s < (p\mathcal{A}_{\mathcal{H}})^{1/2-\delta}$ since $\log \left(1 + \frac{p\mathcal{A}_{\mathcal{H}}}{s^2} \right) \asymp \log p$, meaning the candidate rate can match the minimax rate. Note here we have used that $\mathcal{A}_{\mathcal{H}}$ grows at most logarithmically in p . Therefore, we will take $\mathcal{S} = \{s \in [p] : s \geq \sqrt{p\mathcal{A}_{\mathcal{H}}}\}$ (dropping the subscripts and writing $\mathcal{S} = \mathcal{S}_{p,n}$ for notational ease).

It follows from the fact that $\Gamma_{\mathcal{H}}(s, a)$ is increasing in a that either (17) or (18) should typically hold. To see that $\Gamma_{\mathcal{H}}(s, a)$ is indeed increasing in a , let us fix $a \leq a'$. Consider that for every $\nu \in \mathbb{N}$ we have

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

Taking maximum over ν on both sides yields $\Gamma_{\mathcal{H}}(s, a) \leq \Gamma_{\mathcal{H}}(s, a')$, i.e. $\Gamma_{\mathcal{H}}$ is increasing in a . Now, consider $a = 1$ and $a' = \mathcal{A}_{\mathcal{H}}$ in order to compare the candidate rate ψ_{adapt} to the minimax rate ε^* . Since $s \geq \sqrt{p\mathcal{A}_{\mathcal{H}}}$, we have $\log \left(1 + \frac{p}{s^2} \right) \asymp \frac{p}{s^2}$ and $\log \left(1 + \frac{p\mathcal{A}_{\mathcal{H}}}{s^2} \right) \asymp \frac{p\mathcal{A}_{\mathcal{H}}}{s^2}$. With the above display, it is intuitive that either (17) or (18) should hold in typical nonpathological cases; these conditions can be easily checked on a problem by problem basis (such as those in Section 3.4).

In preparation for the statement of the lower bound, the following set is needed

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

With $\tilde{\mathcal{V}}_{\mathcal{H}}$ defined, we are ready to state the lower bound. The following result establishes condition (i) of Definition 6 for ψ_{adapt} given by (19) with respect to our choice $\mathcal{S} = \{s \in [p] : s \geq \sqrt{p\mathcal{A}_{\mathcal{H}}}\}$.

Theorem 2. Suppose $\mathcal{A}_{\mathcal{H}} \asymp \log(e|\tilde{\mathcal{V}}_{\mathcal{H}}|)$. Assume further $\log(1 + p\mathcal{A}_{\mathcal{H}}) \leq \frac{n}{2}$. 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

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

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

The condition $\mathcal{A}_{\mathcal{H}} \asymp \log(e|\tilde{\mathcal{V}}_{\mathcal{H}}|)$ is somewhat mild; for example it is satisfied when the eigenvalues satisfy a polynomial decay such as $\mu_k = c_{\alpha} k^{-2\alpha}$ for some $\alpha > 0$ and positive universal constant c_{α} . It is also satisfied when there is exponential decay such as $\mu_k = c_1 e^{-c_2 j^{\gamma}}$ for some $\gamma > 0$ and positive universal constants c_1 and c_2 . Sections 3.4 contains further discussion of these two cases.

The proof strategy is, at a high level, the same as that in Section 2.3 with the added complication that the prior should also randomize over the sparsity level in order to capture the added difficulty of not knowing the sparsity level. The random sparsity level is drawn by drawing a random truncation point ν and extracting the induced sparsity. The rest of the prior construction is similar to that in Section 2.3, but the analysis is complicated by the random sparsity level.

Finally, note Theorem 2 is a statement about a cost paid for adaptation for sparsity levels in $\mathcal{S} = \{s \in [p] : s \geq \sqrt{p\mathcal{A}_{\mathcal{H}}}\}$. Nothing is asserted about smaller sparsities. It would be interesting to show whether or not the candidate ψ_{adapt} also captures the cost for sparsity levels s which satisfy $(p\mathcal{A}_{\mathcal{H}})^{1/2-\delta} < s < \sqrt{p\mathcal{A}_{\mathcal{H}}}$ for all $\delta \in (0, 1/2]$ (e.g. $s \asymp \sqrt{p\mathcal{A}_{\mathcal{H}}}/\log(p\mathcal{A}_{\mathcal{H}})$); we leave it open for future work.

3.3 Upper bound

In this section, an adaptive test is constructed. Recall $\mathcal{A}_H$ and $\mathcal{V}_H$ given in (15) and (16) respectively. Define

$$\tau_{\text{adapt}}^2(p, s, n) = \begin{cases} \left(\frac{s}{n} \log \left(1 + \frac{p\mathcal{A}_H}{s^2} \right) \right) \vee \left(\frac{s}{n} \sqrt{\nu_H(s, \mathcal{A}_H) \log \left(1 + \frac{p\mathcal{A}_H}{s^2} \right)} \right) & \text{if } s < \sqrt{p\mathcal{A}_H}, \\ \frac{s}{n} \sqrt{\nu_H(s, \mathcal{A}_H) \log \left(1 + \frac{p\mathcal{A}_H}{s^2} \right)} & \text{if } s \geq \sqrt{p\mathcal{A}_H}. \end{cases} \quad (21)$$

In the test, the following collection of geometric grids is used. Define

$$\mathcal{S} := \left\{ 1, 2, 4, \dots, 2^{\lceil \log_2 \sqrt{p\mathcal{A}_H} \rceil - 1} \right\} \cup \{p\}. \quad (22)$$

The choice to take $\mathcal{S}$ as a geometric grid is not statistically critical, but doing so is computationally advantageous.

Theorem 3. *There exist universal positive constants D, K_2, K'_2 , and K_3 such that the following holds. Let $\mathcal{A}_H$, $\mathcal{V}_H$, and $\mathcal{S}$ be given by (15), (16), and (22) respectively. For $\nu \in \mathcal{V}_H$ and $s \in [p]$, set $d_\nu = \nu \vee \lceil D \rceil$ and set*

$$r_{\nu, s} := K_2 \left(d_\nu \log \left(1 + \frac{p\mathcal{A}_H}{s^2} \right) \right)^{1/4}, \\ r'_s := K'_2 \sqrt{\log \left(1 + \frac{p\mathcal{A}_H}{s^2} \right)}.$$

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

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

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

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

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

As mentioned earlier, the adaptive test involves scanning over the potential values of ν in the geometric grid $\mathcal{V}_H$. Consequently, a cost involving $\mathcal{A}_H$ is paid. Note for $s < \sqrt{p\mathcal{A}_H}$ we also need to scan over $s \in \mathcal{S}$ in order to set the thresholds in the statistics $T_r(d)$ properly. It turns out this extra scan does not incur a cost; cost is driven by needing to scan over ν .

To understand how scanning over ν can incur a cost, it is most intuitive to consider the need to control the Type I error when scanning over the χ^2 statistics with various degrees of freedom. Roughly speaking, it is necessary to pick the threshold values t_ν such that the Type I error $P \left(\bigcup_{\nu \in \mathcal{V}_H} \{ \chi_{\nu p}^2 - \nu p > t_\nu \} \right)$ is smaller than some prescribed error level. By union bound and (14), consider the tail bound $P \left(\bigcup_{\nu \in \mathcal{V}_H} \{ \chi_{\nu p}^2 - \nu p > 2\sqrt{\nu p u} + 2u \} \right) \leq \sum_{\nu \in \mathcal{V}_H} P \{ \chi_{\nu p}^2 - \nu p > 2\sqrt{\nu p u} + 2u \} \leq |\mathcal{V}_H| e^{-u}$. To ensure the Type I error is small, this tail bound suggests the choice $t_\nu \asymp \sqrt{\nu p \log |\mathcal{V}_H|} + 2 \log |\mathcal{V}_H| \asymp \sqrt{\nu p \log |\mathcal{V}_H|}$ where the latter equivalence follows from the fact $|\mathcal{V}_H|$ grows at most logarithmically in p . Since the threshold t_ν is inflated, a signal must also have larger norm in order to be detected. The logarithmic inflation factor is a consequence of the χ^2 tail.

3.4 Special cases

To illustrate how a price for adaptation depends delicately on the choice of function space $\mathcal{H}$, we consider a variety of cases. Throughout, the assumption $\log(1 + p\mathcal{A}_{\mathcal{H}}) \leq n/2$ as well as the other assumptions necessary for our result are made.

3.4.1 Sobolev

Taking $\mu_k \asymp k^{-2\alpha}$ as emblematic of Sobolev space with smoothness α , it can be shown that $\mathcal{A}_{\mathcal{H}} \asymp \log(e|\tilde{\mathcal{Y}}_{\mathcal{H}}|) \asymp \log \log p$ (recall (15) and (20)). Our upper and lower bounds assert an adaptive testing rate is given by

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

Ingster and Lepski [25] consider adaptation in the sparse regime $s = O(p^{1/2-\delta})$ and the dense regime $s = p^{1/2+\delta}$ for $\delta \in (0, 1/2)$ separately. In the sparse regime, they construct an adaptive test which is able to achieve the minimax rate; no cost is paid. In the dense rate, not only do they give a test which achieves $s \left(\frac{ns}{\sqrt{p \log \log p}} \right)^{-\frac{4\alpha}{4\alpha+1}}$, they also supply a lower bound (now, in our lower bound notation, $\mathcal{S} = [p]$ can be taken) showing it cannot be improved. As seen in the above display, our test enjoys optimality in these regimes.

The reader should take care when comparing [25] to our result. In our setting, the unknown sparsity level can vary throughout the entire range $s \in [p]$. In contrast, [25] consider two separate cases. Either the unknown sparsity s is constrained to $s \leq p^{1/2-\delta}$, or it is constrained to $s \geq p^{1/2+\delta}$. To elaborate, the precise value of s is unknown but [25] assumes the statistician has knowledge of which of the two cases is in force. In contrast, we make no such assumption; nothing at all is known about the sparsity level. In our view, this situation is more faithful to the problem facing the practitioner. Despite constructed for a seemingly more difficult setting, it turns out our test is optimal under Ingster and Lepski's setting.

Our result provides finer detail around $\sqrt{p}$ missed by [25]. In Theorem 3 of their article, Ingster and Lepski propose an adaptive test procedure which is applicable in the regime $s \asymp \sqrt{p}$. It requires signal strength of squared order at least

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

This is suboptimal since our test achieves

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

which is faster.

3.4.2 Finite dimension

Consider a finite dimensional structure $1 = \mu_1 = \mu_2 = \dots = \mu_m > \mu_{m+1} = \dots = 0$ where m is a positive integer. If $m < \frac{n^2}{\log(1+p)}$, it can be shown $\mathcal{A}_{\mathcal{H}} \asymp 1$, and so the minimax separation rate can be achieved by our adaptive test. In other words, no cost is paid for adaptation.

3.4.3 Exponential decay

Consider exponential decay of the eigenvalues $\mu_k = c_1 e^{-c_2 k^\gamma}$ where $\gamma > 0$. It can be shown that $\mathcal{A}_{\mathcal{H}} \asymp \log(e|\tilde{\mathcal{V}}_{\mathcal{H}}|) \asymp \log \log \log p$, and so an adaptive rate is

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

The cost for adaptation here grows very slowly, and is perhaps another indication of the relative “small” size of RKHSs based on Gaussian kernels (as noted in Section 2.5).

4 Adaptation to both sparsity and smoothness

So far we have assumed that the space $\mathcal{H}$ is known. However, it is likely that $\mathcal{H}$ is one constituent out of a collection of spaces indexed by some hyperparameter, such as a smoothness level. Typically, the true value of this hyperparameter is unknown in addition to the sparsity being unknown, and it is of interest to understand how the testing rate changes. To avoid excessive abstractness, we follow [17, 25] and adopt a working model where $\mathcal{T}_s$ has μ_k of the form $\mu_k = k^{-2\alpha}$ emblematic of Sobolev space with smoothness α . To emphasize the dependence on α , let us write $\mathcal{T}(s, \alpha)$ and its corresponding space $\mathcal{F}(s, \alpha)$. Reiterating, we are interested in adapting to both the sparsity level s and the smoothness level α .

Ingster and Lepski [25] study adaptation to both sparsity and smoothness. In particular, they study adaptation for an unknown $\alpha \in [\alpha_0, \alpha_1]$ in a closed interval where the endpoints $\alpha_0 < \alpha_1$ are known. As argued in [25], since α_1 is known, the sequence space basis in Section 1.2.2 for any $\alpha \in [\alpha_0, \alpha_1]$ can be taken to be α_1 -regular, and thus is known to the statistician. Furthermore, they separately study the dense regime $s \geq p^{1/2+\delta}$ and the sparse regime $s < p^{1/2-\delta}$ for some constant $\delta \in (0, 1/2)$. We adopt the same setup as them.

Ingster and Lepski [25] claim the following adaptive rate. In the dense case $s \geq p^{1/2+\delta}$, they make the assumption $\frac{p}{s^2} \log\left(\frac{n^2 s^2/p}{\log(n^2 s^2/p)}\right) \rightarrow 0$ and state (Theorem 4 in [25])

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

In contrast¹, we will show the adaptive rate in the dense case $s \geq p^{1/2+\delta}$ is

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

The careful reader will note that in the special case $p = 1$, our answer recovers the adaptive separation rate $(n/\sqrt{\log \log n})^{-\frac{4\alpha}{4\alpha+1}}$ proved by Spokoiny [41]. In the sparse case $s < p^{1/2-\delta}$, their claimed rate (Theorem 5 in [25]) is

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

As a partial correction, we prove the following lower bound

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

¹The same error affects the proofs of the lower bounds in Theorems 4 and 5 in [25]. The prior they define is not supported on the parameter space. Under their prior (see (75) in [25]), the various coordinate functions f_j end up having different smoothness levels α_j instead of sharing a single α .

In the regime $\log p \gtrsim \log \log n$, (23) and (24) match each other and match the minimax rate (see Section 2.5.1). However, in the case $\log p \lesssim \log \log n$, there may be a difference between (23) and (24).

4.1 Dense

Fix $\delta \in (0, 1/2)$ and $\alpha_0 < \alpha_1$. Recall in the dense regime that $s \geq p^{1/2+\delta}$. Define

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

As mentioned earlier, we will be establishing this rate in the dense regime. For use in a testing procedure, define the geometric grid

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

Note the statistician does not need to know s nor α to construct $\mathcal{V}_{\text{test}}$. Further note $\log |\mathcal{V}_{\text{test}}| \asymp \log \log(np)$.

4.1.1 Upper bound

The test procedure employed follows the typical strategy in adaptive testing, namely constructing individual tests for potential values of ν in the grid $\mathcal{V}_{\text{test}}$, and then taking maximum over the individual tests to perform signal detection. The cost of adaptation involves the logarithm of the cardinality of $\mathcal{V}_{\text{test}}$ (i.e. $\log \log(np)$). Since the dense regime is in force, the individual tests are χ^2 tests.

Theorem 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

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

where

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

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

4.1.2 Lower bound

The following theorem states the lower bound. Note that it satisfies Definition 6 with the straightforward modification to incorporate adaptation over α . In particular, the potential difficulty with absurdities outlined in Section 3.2 do not arise since the candidate rate

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

is never of the same order as the minimax rate $\varepsilon^*(p, s, n, \alpha)^2 = s \left(\frac{ns}{\sqrt{p}} \right)^{-\frac{4\alpha}{4\alpha+1}}$.

Theorem 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

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

where τ_{dense} is given by (25).

4.2 Sparse

Fix $\delta \in (0, 1/2)$ and $\alpha_0 < \alpha_1$. Recall that in the sparse regime we have $s < p^{1/2-\delta}$. Define

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

Ingster and Lepski [25] give a testing procedure achieving (23) and thus establish the upper bound. As noted earlier, when $\log p \gtrsim \log \log n$, not only do (27) and (23) match each other but they also match the minimax rate (see Section 2.5.1). In the regime $\log p \lesssim \log \log n$, it is not clear from a lower bound perspective whether a cost for adaptation is unavoidable. We complement their upper bound by providing a lower bound which demonstrates that it is indeed necessary to pay a cost for adaptation. However, the cost we identify may not be sharp. To elaborate, in the case where p is a large universal constant but much smaller than n , the sparse regime rate (27) involves $\sqrt{\log \log \log n}$. From Spokoiny's article [41], $\sqrt{\log \log n}$ is instead expected. We leave it for future work to pin down the sharp cost.

Theorem 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

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

where τ_{sparse} is given by (27).

As discussed with respect to Theorem 5, the lower bound in Theorem 6 satisfies Definition 6 with the straightforward modification to incorporate adaptation over α .

5 Discussion

5.1 Relaxing the centered assumption

In Section 1.2.2, in the definition of the parameter space (4) the signal f was constrained to be centered. It was claimed this constraint is mild and can be relaxed; this section elaborates on this point.

Suppose f is an additive function (possibly uncentered) with each $f_j \in \mathcal{H}$. Let $\bar{f} = \int_{[0,1]^p} f(x) dx$ and note we can write $f = \bar{f} \mathbf{1}_p + (f - \bar{f} \mathbf{1}_p)$ where $\mathbf{1}_p \in L^2([0,1]^p)$ is the constant function equal to one. Since f is an additive function, for any $x \in [0,1]^p$ we have $f(x) = \bar{f} + g(x)$ where $g(x) = \sum_{j=1}^p g_j(x_j)$ with $g_j(x_j) = f_j(x_j) - \int_0^1 f_j(t) dt$. In other words, we have $f = \bar{f} \mathbf{1}_p + g$ where g itself is an additive function with each $g_j \in \mathcal{H}_0$.

Asserting f is a sparse additive function implies g is a (centered) sparse additive function. Following [17, 38], we impose constraints (i.e. RKHS norm constraints) on the centered component functions g_j . Therefore, the following uncentered signal detection problem can be considered,

$$H_0 : f \equiv 0, \quad (28)$$

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

where $\mathcal{F}_s$ is given by (4). Let ε^* denote the minimax separation rate. Consider $\|f\|_2^2 = \bar{f}^2 + \|g\|_2^2$ by orthogonality. Thus, there are two related detection problems, Problem I

$$H_0 : f \equiv 0,$$

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

and Problem II

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

Let ε_1^* and ε_2^* denote the minimax separation rates of the respective problems. We claim $(\varepsilon^*)^2 \asymp (\varepsilon_1^*)^2 + (\varepsilon_2^*)^2$. From $\|f\|^2 = \bar{f}^2 + \|g\|_2^2 \geq \bar{f}^2 \vee \|g\|_2^2$ it is directly seen that $(\varepsilon^*)^2 \gtrsim (\varepsilon_1^*)^2 \vee (\varepsilon_2^*)^2 \asymp (\varepsilon_1^*)^2 + (\varepsilon_2^*)^2$, and so the lower bound is proved. To show the upper bound, consider that if $\|f\|_2^2 \geq C((\varepsilon_1^*)^2 + (\varepsilon_2^*)^2)$, then by triangle inequality it follows that either $|\bar{f}|^2 \geq C(\varepsilon_1^*)^2$ or $\|g\|_2^2 \geq C(\varepsilon_2^*)^2$. Thus, there is enough signal to detect and the test which takes maximum over the subproblem minimax tests is optimal. Hence, the upper bound $(\varepsilon^*)^2 \lesssim (\varepsilon_1^*)^2 + (\varepsilon_2^*)^2$ is now also proved.

All that remains is to determine the minimax rates ε_1^* and ε_2^* . Let us index starting from zero and take $\{\psi_k\}_{k=0}^\infty$ and $\{\mu_k\}_{k=0}^\infty$ to be the associated eigenfunctions and eigenvalues of the RKHS $\mathcal{H}_0$. Recall $\{\psi_k\}_{k=0}^\infty$ forms an orthonormal basis for $L^2([0, 1])$ under the usual L^2 inner product. Since by definition $\mathcal{H}_0$ is orthogonal to the span of the constant function in $L^2([0, 1])$, without loss of generality we can take ψ_0 to be the constant function equal to one, $\mu_0 = 0$, and $\{\psi_k\}_{k=1}^\infty$ to be the remaining eigenfunctions orthogonal (in L^2) to constant functions.

The data $\{X_{k,j}\}_{k \in \mathbb{N}, j \in [p]}$ (defined in (5)) is sufficient for Problem II. Therefore, the minimax rate ε_2^* is exactly given by our result established in this paper. The minimax rate ε_1^* can be upper bounded in the following manner. Consider that $\langle \mathbf{1}_p, dY_x \rangle_{L^2([0, 1]^p)} \sim N(\bar{f}, \frac{1}{n})$ since $\|\mathbf{1}_p\|_2^2 = 1$. Therefore, the parametric rate $\varepsilon_1^* \lesssim \frac{1}{\sqrt{n}}$ can be achieved. Hence, $(\varepsilon^*)^2 \asymp (\varepsilon_1^*)^2 + (\varepsilon_2^*)^2 \asymp (\varepsilon_2^*)^2$. In other words, centering is immaterial to the fundamental limits of the signal detection problem.

5.2 Sharp constant

As we noted earlier, our results do not say anything about the sharp constant in the detection boundary. The problem of obtaining a sharp characterization of the constants in the detection boundary is interesting and likely delicate. In an asymptotic setup and in the Sobolev case, Gayraud and Ingster [17] were able to derive the sharp constant in the sparse regime $s = p^{1-\beta}$ for fixed $\beta \in (1/2, 1)$ under the condition $\log p = o(n^{\frac{1}{2\alpha+1}})$. Gayraud and Ingster discuss that this condition is actually essential, in that the detection boundary no longer exhibits a dependence on β when the condition is violated. This condition has a nice formulation in our notation, namely it is equivalent to the condition

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

This correspondence in the Sobolev case suggests this condition may actually be essential for a generic $\mathcal{H}$. It would be interesting to understand if that is true, and to derive the sharp constant when it holds if it so.

To be clear, it may be the case (perhaps likely) our proposed procedure is suboptimal in terms of the constant. Indeed, the existing literature on sparse signal detection, both in sparse sequence model [14, 18] and sparse linear regression [1, 26] rely on Higher Criticism type tests to achieve the optimal constant. Gayraud and Ingster [17] themselves use Higher Criticism. For a generic space $\mathcal{H}$, our procedure should not be understood as the only test which is rate-optimal. In the sparsity regime $s = p^{1/2-\delta}$, we suspect an analogous Higher Criticism type statistic which accounts for the eigenstructure of the kernel might not only achieve the optimal rate, but also the sharp constant.

5.3 Future directions

There are a number of avenues for future work. First, we only considered one space $\mathcal{H}$ in which all component functions f_j live in. In some scenarios, it may be desirable to consider a different space $\mathcal{H}_j$ for each component. Raskutti et al. [38] obtained the minimax estimation rate when considering multiple kernels (under some conditions on the kernels). We imagine our broad approach in this work could be extended to determine the minimax separation rate allowing multiple kernels. Instead of a

common $\nu_{\mathcal{H}}$, it is likely different quantities $\nu_{\mathcal{H}_j}$ will be needed per coordinate and the test statistics could be modified in the natural way. The theory developed here could be used directly, and it seems plausible the minimax separation rate could be established in a straightforward manner.

Another avenue of research involves considering “soft” sparsity in the form of ℓ_q constraints for $0 < q < 1$. Yuan and Zhou [50] developed minimax estimation rates for RKHSs exhibiting polynomial decay of its eigenvalues (e.g. Sobolev space). In terms of signal detection, it is plausible that the quadratic functional estimator under the Gaussian sequence model with ℓ_q bound on the mean could be extended and used as a test statistic [10]. The hard sparsity and soft sparsity settings studied in [10] are handled quite similarly. It is possible not much additional theory needs to be developed in order to obtain minimax separation rates under soft sparsity.

Since hypothesis testing is quite closely related to functional estimation in many problems, it is natural to ask about functional estimation in the context of sparse additive models. For example, it would be of interest to estimate the quadratic functional $\|f\|_2^2$ or the norm $\|f\|_2$. It is well known in the nonparametric literature that estimating L_r norms for odd r yields drastically different rates from testing and from even r [19, 32]. A compelling direction is to investigate the same problem in the sparse additive model setting.

Additionally, it would be interesting to consider the nonparametric regression model $Y_i = f(X_i) + Z_i$ where the design distribution $X_i \stackrel{iid}{\sim} P_X$ exhibits some dependence between the coordinates. The correspondence between the white noise model (1) and the nonparametric regression model we relied on requires that the design distribution be close to uniform on $[0, 1]^p$. However, in practical situations it is typically the case that the coordinates of X_i exhibit dependence, and it would be interesting to understand how the fundamental limits of testing are affected.

Finally, in a somewhat different direction than that discussed so far, it is of interest to study the signal detection problem under group sparsity in other models. As we had encouraged, our results can be interpreted exclusively in terms of sequence space, that is the problem (7)-(8) with parameter space (6). From this perspective, the group sparse structure is immediately apparent. Estimation has been extensively studied in group sparse settings, especially in linear regression (see [47] and references therein). Hypothesis testing has not witnessed nearly the same level of research activity, and so there is ample opportunity. We imagine some features of the rates established in this paper are general features of the group sparse structure, and it would be intriguing to discover the commonalities.

6 Acknowledgments

SK is grateful to Oleg Lepski for a kind correspondence. The research of SK is supported in part by NSF Grants DMS-1547396 and ECCS-2216912. The research of CG is supported in part by NSF Grant ECCS-2216912, NSF Career Award DMS-1847590, and an Alfred Sloan fellowship.

7 Proofs

7.1 Minimax upper bounds

7.1.1 Sparse

Proof of Proposition 2. Fix $\eta \in (0, 1)$. We will set C_η at the end of the proof, so for now let $C > C_\eta$. The universal constant K_1 will be selected in the course of the proof. Let L^* denote the universal constant from Lemma 21. Set $K_2 := 1 \vee \frac{1}{(\log 2)^{1/4}} \vee c^{-1/4} \vee \left(\frac{L^*}{c}\right)^{1/4}$ and $K_3 := \frac{L^*}{K_2^2}$ where $c = c^* \wedge c^{**}$ where c^* and c^{**} are the universal constants in the exponential terms of Lemmas 15 and 14 respectively.

We first bound the Type I error. Since $\sqrt{\log(1 + \frac{p}{s^2})} \leq K_3 \sqrt{d}$, we have $1 \leq K_2^2 \sqrt{\log(1 + \frac{p}{s^2})} \leq$

$K_2^2 K_3 \sqrt{d} \leq L^* \sqrt{d}$. Therefore, we can apply Lemma 15 to obtain that

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

for any $x > 0$. Here, C^* is a universal constant. Taking $x = C$ and noting that $C > 1$ provided we select $C_\eta \geq 1$, we see that

$$\begin{aligned} C^* \left(\sqrt{xpr^4 e^{-\frac{c^* r^4}{d}}} + \frac{d}{r^2} x \right) &\leq C^* C \left(K_2^2 \sqrt{pd \log \left(1 + \frac{p}{s^2} \right)} e^{-c^* K_2^4 \log \left(1 + \frac{p}{s^2} \right)} + \frac{d}{r^2} \right) \\ &\leq C^* C \left(K_2^2 \sqrt{d \log \left(1 + \frac{p}{s^2} \right)} \sqrt{p \cdot \frac{s^2}{s^2 + p}} + \frac{d}{r^2} \right) \\ &\leq C^* C \left(K_2^2 s \sqrt{d \log \left(1 + \frac{p}{s^2} \right)} + \frac{d}{r^2} \right) \\ &\leq 2C^* K_2^2 C s \sqrt{d \log \left(1 + \frac{p}{s^2} \right)} \\ &= CK_1 n \tau(p, s, n)^2 \end{aligned}$$

where we have used that $c^* K_2^4 \geq 1$ and where we have selected $K_1 = 2C^* K_2^2$. Note we have also used that $\frac{d}{r^2} \leq \sqrt{d} \leq K_2^2 s \sqrt{d \log \left(1 + \frac{p}{s^2} \right)}$. Thus, with these choices of K_1, K_2 , and x , we have

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

provided we select $C_\eta \geq \log \left(\frac{2}{\eta} \right) \vee 1$.

We now examine the Type II error. To bound the Type II error, we will use Chebyshev's inequality. In particular, consider that for any $f \in \mathcal{F}_s$ with $\|f\|_2 \geq C\tau(p, s, n)$, we have

$$\begin{aligned} P_f \{ T_r(d) \leq CK_1 n \tau(p, s, n)^2 \} &= P_f \{ E_f(T_r(d)) - CK_1 n \tau(p, s, n)^2 \leq E_f(T_r(d)) - T_r(d) \} \\ &\leq \frac{\text{Var}_f(T_r(d))}{(E_f(T_r(d)) - CK_1 n \tau(p, s, n)^2)^2} \end{aligned} \quad (30)$$

provided that $E_f(T_r(d)) \geq CK_1 n \tau(p, s, n)^2$. To ensure this application of Chebyshev's inequality is valid, we must compute suitable bounds for the expectation and variance of $T_r(d)$, which the following lemmas provide.

Lemma 4. *If C_η is larger than some sufficiently large universal constant, then $E_f(T_r(d)) \geq 2CK_1 n \tau(p, s, n)^2$.*

Lemma 5. *If C_η is larger than some sufficiently large universal constant, then $\text{Var}_f(T_r(d)) \leq C^\dagger (s^2 r^4 + E_f(T_r(d)))$ where $C^\dagger > 0$ is a universal constant.*

These bounds are proved later on. Let us now describe why they allow us to bound the Type II

error. From (30) as well as Lemmas 4 and 5, we have

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

provided we pick C_η larger than a sufficiently large universal constant as required by Lemmas 4 and 5. With this bound in hand and since $C \geq C_\eta$, we can now pick C_η sufficiently large depending only on η to obtain

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

To summarize, we can pick C_η depending only on η to be sufficiently large and also satisfying the condition $C_\eta \geq \log\left(\frac{2}{\eta}\right) \vee 1$ and those of Lemmas 4, 5 to ensure the testing risk is bounded by η , i.e.

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

as desired. $\square$

It remains to prove Lemmas 4 and 5. Recall we work in the environment of the proof of Proposition 2.

Proof of Lemma 4. To prove the lower bound on the expectation, first recall

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

Under P_f , we have $E_j(d) \sim \chi_d^2(m_j^2)$ where $m_j^2 = n \sum_{k \leq d} \theta_{k,j}^2$. Here, the collection $\{\theta_{k,j}\}$ denotes the basis coefficients of f . Intuitively, there are two reasons why the expectation might be small. First, we are thresholding and so we are intuitively removing those coordinates with small, but nonetheless nonzero, means. Furthermore, we are truncating at level d and not considering higher-order basis coefficients; this also incurs a loss in signal.

Let us first focus on the effect from thresholding. Let $\tilde{C}$ denote the universal constant from Lemma

13. Applying Lemma 13 yields

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

where $S_f \subset [p]$ denotes the subset of active variables j such that $\Theta_j \neq 0$. Note that such S_f exists and $|S_f| \leq s$ since $f \in \mathcal{F}_s$. We have thus bounded the amount of signal lost from thresholding.

Let us now examine the effect of truncation. Consider that

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

Here, we have used $\sum_{k=1}^{\infty} \mu_k^{-1} \theta_{k,j}^2 \leq 1$ for all j . Therefore, we have shown $\sum_{j \in S_f} \sum_{k \leq d} \theta_{k,j}^2 \geq \|f\|_2^2 - s\mu_{d+1}$, and thus have quantified the loss due to truncation.

We are now in position to put together the two pieces. Consider $\|f\|_2^2 \geq C^2 \tau(p, s, n)^2$. Consequently,

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

We have used the decreasing order of the kernel's eigenvalues, i.e. that $d \geq \nu_{\mathcal{H}}$ implies $\mu_{\nu_{\mathcal{H}}+1} \geq \mu_{d+1}$. Further, consider that by definition of $\nu_{\mathcal{H}}$ and $\tau(p, s, n)^2$, we have

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

With this in hand, it follows that $\sum_{j \in S_f} \frac{m_j^2}{2} \geq \left(\frac{C^2-1}{2} \right) n \tau(p, s, n)^2$. To summarize, we have shown

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

Here, we have used $d \leq \lceil D \rceil \nu_{\mathcal{H}}$. We can also conclude

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

In view of the above bound and since $C > C_\eta$, it suffices to pick C_η large enough to satisfy $C_\eta^2 - 4K_1C_\eta \geq 1 + K_2^2 \tilde{C}$ to ensure $E_f(T_r(d)) \geq 2CK_1n\tau(p, s, n)^2$. The proof of the lemma is complete. $\square$

Proof of Lemma 5. To bound the variance of $T_r(d)$, recall that $\sqrt{\log(1 + \frac{p}{s^2})} \leq K_3\sqrt{d}$, and so $1 \leq K_2^2 \sqrt{\log(1 + \frac{p}{s^2})} \leq K_2^2 K_3 \sqrt{d} \leq L^* \sqrt{d}$. Therefore, we can apply Lemma 14. By Lemma 14, we have

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

where $C^\dagger$ is a positive universal constant whose value can change from instance to instance. Recall c^{**} is defined at the beginning of the proof of Proposition 2. Since $r^4 = K_2^4 d \log(1 + \frac{p}{s^2})$ and $c^{**}K_2^4 \geq 1$, we have

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

Therefore,

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

Taking C_η larger than a sufficiently large universal constant, noting $C \geq C_\eta$, and invoking (31), we have the desired result. $\square$

Proof of Proposition 3. Our proof is largely the same as in the proof of Proposition 2, except we invoke results about the “tail” rather than the “bulk”. Fix $\eta \in (0, 1)$. We will make a choice of C_η at the end of the proof, so for now let $C > C_\eta$. We select K_1 in the course of the proof, but we will select K_2 now. Set $K_2 := \frac{1}{\sqrt{\log 2}} \vee c^{-1/2} \vee (cK_3^2)^{-1/4}$ with $c = c^* \wedge c^{**}$ where c^* and c^{**} are the universal constants in the exponential terms of Lemmas 11 and 10 respectively.

We first bound the Type I error. Since $\log(1 + \frac{p}{s^2}) \geq K_2^2 d$, we have $r^2 \gtrsim d$. Since $d \geq D$, we have by Lemma 11 that for any $x > 0$,

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

where C^* is a universal positive constant. Taking $x = C$ and noting $C > 1$ provided we have chosen $C_\eta \geq 1$, we see that

$$\begin{aligned} C^* \left(\sqrt{xpr^4 e^{-c^*r^2}} + x \right) &\leq C^* C \left(K_2^2 \log \left(1 + \frac{p}{s^2} \right) \sqrt{pe^{-c^*K_2^2 \log(1 + \frac{p}{s^2})}} + 1 \right) \\ &\leq C^* C \left(K_2^2 \log \left(1 + \frac{p}{s^2} \right) \sqrt{p \cdot \frac{s^2}{s^2 + p}} + 1 \right) \\ &\leq 2C^* CK_2^2 s \log \left(1 + \frac{p}{s^2} \right) \\ &= CK_1 n\tau(p, s, n)^2 \end{aligned}$$

where we have used that $c^*K_2^2 \geq 1$ and we have set $K_1 := 2C^*K_2^2$. Thus, with these choices of

K_1, K_2 , and x we have

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

provided we select $C_\eta \geq \log\left(\frac{2}{\eta}\right) \vee 1$.

We now examine the Type II error. To bound the Type II error, we will use Chebyshev's inequality. In particular, consider that for any $f \in \mathcal{F}_s$ with $\|f\|_2 \geq C\tau(p, s, n)$, we have

$$\begin{aligned} P_f \{T_r(d) \leq CK_1 n \tau(p, s, n)^2\} &= P_f \{E_f(T_r(d)) - CK_1 n \tau(p, s, n)^2 \leq E_f(T_r(d)) - T_r(d)\} \\ &\leq \frac{\text{Var}_f(T_r(d))}{(E_f(T_r(d)) - CK_1 n \tau(p, s, n)^2)^2} \end{aligned} \quad (32)$$

provided that $E_f(T_r(d)) \geq CK_1 n \tau(p, s, n)^2$. To ensure this application of Chebyshev's inequality is valid and to bound the Type II error, we will need a lower bound on the expectation of $T_r(d)$. We will also need an upper bound on the variance of $T_r(d)$ in order to bound the Type II error. The following lemmas provide us with the requisite bounds; they are analogous to Lemmas 4 and 5 but are now in the context of the tail regime.

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

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

With these bounds in hand, the argument in the proof of Proposition 2 can be essentially repeated to establish that

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

provided $C_\eta \geq \log\left(\frac{2}{\eta}\right) \vee 1$ and C_η sufficiently large to satisfy Lemmas 6 and 7. We omit the details for brevity. $\square$

It remains to prove Lemmas 6 and 7.

Proof of Lemma 6. The proof is similar in style to the proof of Lemma 4, except now results for the tail regime are invoked. Letting $\tilde{C}$ denote the universal constant from Lemma 9, applying Lemma 9, and arguing similarly to the proof of Lemma 4, we obtain

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

where $S_f \subset [p]$ denotes the subset of active variables j such that $\Theta_j \neq 0$. Note that such S_f exists and $|S_f| \leq s$ since $f \in \mathcal{F}_s$. Further arguing like the proof of Lemma 4 and using $\sqrt{\log\left(1 + \frac{p}{s^2}\right)} \geq K_3 \sqrt{d}$, we have

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

To summarize, we have shown

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

Note that we also can conclude

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

Since $C > C_\eta$, it suffices to pick C_η large enough to satisfy $C_\eta^2 - 4K_1C_\eta \geq \frac{1}{K_3} + K_2^2 \tilde{C}$ to ensure $E_f(T_r(d)) \geq 2CK_1 n \tau(p, s, n)^2$. The proof of the lemma is complete. $\square$

Proof of Lemma 7. Recall that $\log(1 + \frac{p}{s^2}) \geq K_3^2 d$. By definition of r^2 we have $r^2 \geq K_2^2 K_3^2 d$. In other words $r^2 \gtrsim d$ and so we can apply Lemma 10. By Lemma 10, we have

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

where $C^\dagger$ is a positive universal constant. Recall we had defined c^{**} at the beginning of the proof of Proposition 3. Since $r^2 \geq K_2^2 K_3^2 d$, we have $\frac{r^4}{d} \geq r^2 K_2^2 K_3^2$. Therefore,

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

where we have used that $c^{**} (K_2^2 K_3^2 \wedge 1) K_2^2 \geq 1$ by definition of K_2 . Therefore,

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

Taking C_η larger than a sufficiently large universal constant and invoking (34) yields the desired result. $\square$

7.1.2 Dense

Proof of Proposition 4. For ease of notation, set $L = \frac{2}{\sqrt{\eta}}$. Define $C_\eta := \left(\sqrt{2\sqrt{2}L+1} \right) \vee \left(\sqrt{1 + \frac{4}{\sqrt{\eta}}} + L \right) \vee \left(\sqrt{\frac{64\sqrt{2}}{\eta}} + 1 \right)$. Let $C > C_\eta$. We first bound the Type I error. Consider that under P_0 we have $T \sim \chi_{p\nu_{\mathcal{H}}}^2$. Observe that $E_0(T) = p\nu_{\mathcal{H}}$ and $\text{Var}_0(T) = 2p\nu_{\mathcal{H}}$. Consequently, we have by Chebyshev's inequality

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

We now examine the Type II error. Under P_f , we have $T \sim \chi_{p\nu_{\mathcal{H}}}^2 \left(n \sum_{j=1}^p \sum_{k \leq \nu_{\mathcal{H}}} \theta_{k,j}^2 \right)$ where $\{\theta_{k,j}\}$ denote the basis coefficients associated to f . Note $E_f(T) = p\nu_{\mathcal{H}} + n \sum_{j=1}^p \sum_{k \leq \nu_{\mathcal{H}}} \theta_{k,j}^2$ and $\text{Var}_f(T) = 2p\nu_{\mathcal{H}} + 4n \sum_{j=1}^p \sum_{k \leq \nu_{\mathcal{H}}} \theta_{k,j}^2$. In order to proceed with the argument, we need to obtain a lower bound estimate for the signal strength. Letting S denote the set of j for which Θ_j are nonzero,

consider that for $\|f\|_2^2 \geq C^2 \tau(p, s, n)^2$ we have

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

We have used $\sum_{k=1}^{\infty} \frac{\theta_{k,j}^2}{\mu_k} \leq 1$ for all j in the final line. Therefore by definition of $\nu_{\mathcal{H}}$ we have

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

where we have used that $n s \mu_{\nu_{\mathcal{H}}} \leq s \sqrt{\nu_{\mathcal{H}} \log(1 + \frac{p}{s^2})} \leq \sqrt{p \nu_{\mathcal{H}}}$.

We now continue with bounding the Type II error. By Chebyshev's inequality, we have

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

Therefore, the sum of Type I and Type II errors is bounded by η as desired. $\square$

7.2 Minimax lower bounds

Proof of Proposition 1. We break up the analysis into two cases.

Case 1: Suppose $s < \sqrt{p}$. We will construct a prior distribution π on $\mathcal{T}_s$ and use Le Cam's two

point method to furnish a lower bound. Define $c_\eta := 1 \wedge \sqrt{\kappa} \wedge \sqrt{\kappa \log(1 + 4\eta^2)}$. Let $0 < c < c_\eta$. Let π be the prior in which a draw $\Theta \sim \pi$ is constructed by uniformly drawing $S \subset [p]$ of size s and setting

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

where $e_1 \in \mathbb{R}^N$ is given by $e_1 = (1, 0, 0, \dots)$. Note that $\Theta \sim \pi$ implies $\|\Theta\|_F^2 = c^2 s$ and $\Theta \in \mathcal{T}_s$. By Neyman-Pearson lemma and the inequality $1 - d_{TV}(Q, P) \geq 1 - \sqrt{\chi^2(Q||P)}/2$, we have

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

where P_π denotes the mixture $\int P_\Theta \pi(d\Theta)$ induced by π . By the Ingster-Suslina method (Proposition 5) and Corollary 3, we have

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

where $\Theta, \tilde{\Theta} \stackrel{iid}{\sim} \pi$ and $S, \tilde{S}$ are the corresponding random sets. Now, using that $\log(1 + \frac{p}{s^2}) \geq \kappa n$, we have

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

We have also used that $c^2 \kappa^{-1} < c_\eta^2 \kappa^{-1} \leq 1$ and the inequality $(1 + x)^y \leq 1 + xy$ for $x > 0$ and $y \in (0, 1)$. Plugging into (35) yields the desired result.

Case 2: Suppose $s \geq \sqrt{p}$. We can repeat the analysis done in Case 1 with the modification of replacing every instance of s with $\lceil \sqrt{p} \rceil$. $\square$

Proof of Theorem 1. We will construct a prior distribution π on $\mathcal{T}_s$ and use Le Cam's two point method to furnish a lower bound. Set $c_\eta := 1 \wedge 2^{1/4} \wedge (\log(1 + 4\eta^2))^{1/4}$. Let $0 < c < c_\eta$. We break up the analysis into two cases.

Case 1: Suppose we are in the regime where $\psi(p, s, n)^2 = s\Gamma_{\mathcal{H}}$. Set $\rho := \sqrt{\frac{\Gamma_{\mathcal{H}}}{\nu_{\mathcal{H}}}}$. Let π be the prior in which a draw $\Theta \sim \pi$ is obtained by uniformly drawing $S \subset [p]$ of size s and drawing

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

where δ_0 denotes the probability measure placing full mass at zero. Note that $\Theta \sim \pi$ implies $\|\Theta\|_F^2 = c^2 \rho^2 s(\nu_{\mathcal{H}} - 1) = c^2 s \Gamma_{\mathcal{H}} \cdot \frac{\nu_{\mathcal{H}} - 1}{\nu_{\mathcal{H}}} \geq \frac{c^2}{2} s \Gamma_{\mathcal{H}}$. Here, we have used that $\nu_{\mathcal{H}} \geq 2$ since $\log(1 + \frac{p}{s^2}) \leq \frac{n}{2}$. Furthermore, consider that for $i \in S$, we have

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

Here, we have used the ordering of the eigenvalues $\mu_1 \geq \mu_2 \geq \dots \geq 0$ and $\Gamma_{\mathcal{H}} \leq \mu_{\nu_{\mathcal{H}}-1}$ by Lemma 25. Of course, for $i \notin S$ we have $\Theta_i = 0$. Hence we have $\Theta \in \mathcal{T}_s$. We then have by the Neyman-Pearson lemma and the inequality $1 - d_{TV}(Q, P) \geq 1 - \sqrt{\chi^2(Q||P)}/2$ that

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

For the following calculations, let $\Theta, \Theta' \stackrel{iid}{\sim} \pi$ and S, S' are the corresponding random sets. Also, let $\{r_{i\ell}, r'_{i\ell}\}_{1 \leq i \leq p, \ell \in \mathbb{N}}$ denote an iid collection of Rademacher (1/2) random variables. By the Ingster-Suslina method (Proposition 5), independence of $\{S, S'\}$ with $\{r_{i\ell}, r'_{i\ell}\}_{1 \leq i \leq p, \ell \in \mathbb{N}}$, and Corollary 3, we have

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

Consider by Lemma 1 that $c^4 n^2 \rho^4 \nu_{\mathcal{H}} = \frac{c^4 n^2 \Gamma_{\mathcal{H}}^2}{\nu_{\mathcal{H}}} \leq c^4 \log \left(1 + \frac{p}{s^2} \right)$. Therefore,

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

We have used $\frac{c^4}{2} < \frac{c^4}{2} \leq 1$ and the inequality $(1+x)^y \leq 1+xy$ for $x > 0$ and $y \in (0, 1)$. Using $\chi^2(P_{\pi}||P_0) \leq 4\eta^2$ with (36) yields the desired result.

Case 2: Suppose $s < \sqrt{p}$ and $\psi(p, s, n)^2 = \frac{s}{n} \log \left(1 + \frac{p}{s^2} \right)$. Set $\rho = 1$. Let π be the prior in which a draw $\Theta \sim \pi$ is obtained by uniformly drawing $S \subset [p]$ of size s and drawing

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

The desired result can be proved by arguing in a manner similar to that as in the proof of Theorem 1 (see also [10]). Details are omitted for the sake of brevity. $\square$

7.3 Adaptive lower bound

Proof of Theorem 2. Recall the definitions of $\mathcal{A}_{\mathcal{H}}$ in (15) and $\tilde{\mathcal{V}}_{\mathcal{H}}$ in (20). By assumption, we have $|\mathcal{A}_{\mathcal{H}}| \leq L \log(e|\tilde{\mathcal{V}}_{\mathcal{H}}|)$ for a universal constant $L > 0$. For ease of notation, let us denote $\psi = \psi_{\text{adapt}}(p, s, n)$ where ψ_{adapt} is given by (19). Set $c_{\eta} := (16L)^{-1/4} \wedge \left(\frac{\eta^2 \log(1+2\eta^2)}{32L} \right)^{1/4}$ and let $0 < c < c_{\eta}$. We now define a prior π . A draw $\Theta \sim \pi$ is obtained as follows. First, draw $v \sim \text{Uniform}(\tilde{\mathcal{V}}_{\mathcal{H}})$. Set $s := \min \{s' \in [p] : s' \geq \sqrt{p|\mathcal{A}_{\mathcal{H}}|} \text{ and } \frac{v}{2} < \nu_{\mathcal{H}}(s', \mathcal{A}_{\mathcal{H}}) \leq v\}$. Then draw uniformly at random a subset $S \subset [p]$ of size exactly s . Let $d_s := 2^{k_s} \in \tilde{\mathcal{V}}_{\mathcal{H}}$ satisfy $2^{k_s-1} < \nu_{\mathcal{H}}(s, \mathcal{A}_{\mathcal{H}}) \leq 2^{k_s}$.

Draw independently

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

Here, $\rho_s = \sqrt{\frac{\psi^2/s}{\nu_{\mathcal{H}}(s, \mathcal{A}_{\mathcal{H}})}}$. This description concludes the definition of π .

Now we must show that π is indeed supported on $\bigcup_s \mathcal{T}_s$. Note that when we draw $\Theta \sim \pi$, the associated sparsity level s always satisfies $s \geq \sqrt{p_{\mathcal{A}_{\mathcal{H}}}}$. Furthermore, conditional on s we have

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

Furthermore, consider that for $j \in S$ we have

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

where the last line follows from $c < c_{\eta}$. Note we have used the ordering of the eigenvalues as well as $\Gamma_{\mathcal{H}}(s, \mathcal{A}_{\mathcal{H}}) \leq \mu_{\nu_{\mathcal{H}}(s, \mathcal{A}_{\mathcal{H}})-1}$ by Lemma 26. Of course, for $j \notin S$ we have $\Theta_j = 0$. Hence, $\Theta \in \mathcal{T}_s$ and so π is properly supported.

Writing $P_{\pi} = \int P_{\Theta} \pi(d\Theta)$ for the mixture, we have

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

For the following calculations, let $\Theta, \Theta' \stackrel{iid}{\sim} \pi$. Let $v, v' \in \tilde{\mathcal{V}}_{\mathcal{H}}$ be the corresponding quantities. Let s and t denote the corresponding sparsities. Denote the corresponding support sets S and T . Also, let $\{r_{i\ell}, \tilde{r}_{i\ell}\}_{1 \leq i \leq p, \ell \in \mathbb{N}}$ denote an iid collection of Rademacher $(1/2)$ random variables which is independent of s, t, S , and T . For ease of notation, let $\nu_s = \nu_{\mathcal{H}}(s, \mathcal{A}_{\mathcal{H}})$. Likewise, let $\nu_t = \nu_{\mathcal{H}}(t, \mathcal{A}_{\mathcal{H}})$. By the Ingster-Suslina method (Proposition 5), we have

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

Here, we have used the inequality $\cosh(x) \leq e^{x^2/2}$ for $x > 0$. Consider that by Lemma 26

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

We have used $s, t \geq \sqrt{p_{\mathcal{A}_{\mathcal{H}}}}$ along with the inequality $\frac{u}{2} \leq \log(1+u) \leq u$ for $0 \leq u \leq 1$. Therefore,

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

Since $8c^4 \frac{d_s \wedge d_t}{\sqrt{d_s d_t}} \leq 1$, we can use Lemma 28 and the inequality $(1+x)^\delta \leq 1 + \delta x$ for $x > 0$ and $\delta \leq 1$ to obtain

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

Recall we write $d_s = 2^{k_s}$ and $d_t = 2^{k_t}$. Moreover, recall $d_s, d_t \in \tilde{\mathcal{V}}_{\mathcal{H}}$. Now observe

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

Recall we have $\mathcal{A}_{\mathcal{H}} \leq L \log(e|\tilde{\mathcal{V}}_{\mathcal{H}}|)$. Define the sets

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

Then

$$\begin{aligned}
&E\left(\exp\left(8c^4 2^{-\frac{|k_s - k_t|}{2}} \mathcal{A}_{\mathcal{H}}\right)\right) \\
&= E\left(\exp\left(8c^4 2^{-\frac{|k_s - k_t|}{2}} \mathcal{A}_{\mathcal{H}}\right) \mathbb{1}_{\{(k_s, k_t) \in E_1\}}\right) + E\left(\exp\left(8c^4 2^{-\frac{|k_s - k_t|}{2}} \mathcal{A}_{\mathcal{H}}\right) \mathbb{1}_{\{(k_s, k_t) \in E_2\}}\right).
\end{aligned}$$

Let us examine the second term. Note that for any $\ell > 0$ we have $x^{-\ell} \log(x) \leq (e\ell)^{-1}$ for all $x > 0$.

Using this, we obtain

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

The final line results from $c^4 < c_{\eta}^4 \leq \frac{1}{32L}\eta^2 \log(1 + 2\eta^2)$. Note we have also used $8Lc^4 - 1 \leq 0$ to obtain the penultimate line. We now examine E_1 . Consider that $|E_1| \leq \frac{\eta^2}{2e^{8Lc^4}} |\tilde{\mathcal{V}}_{\mathcal{H}}| \log_2(e|\tilde{\mathcal{V}}_{\mathcal{H}}|)$. Therefore,

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

Note we have used $\mathcal{A}_{\mathcal{H}} \leq L \log(e|\tilde{\mathcal{V}}_{\mathcal{H}}|)$. We have also used $8Lc^4 < 8Lc_{\eta}^4 \leq \frac{1}{2}$ along with the inequality $\log(x) \leq \sqrt{x}$ for all $x > 0$ to obtain the penultimate line. Therefore, we have shown

$$\chi^2(P_{\pi} || P_0) \leq 2\eta^2 + (2\eta^2 + 1) - 1 = 4\eta^2.$$

Plugging into (37) yields

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

The proof is complete. $\square$

7.4 Adaptive upper bound

Proof of Theorem 3. Fix $\eta \in (0, 1)$. In various places of the proof, we will point out that C_{η} can be taken sufficiently large to obtain desired bounds, so for now let $C > C_{\eta}$. Let L^* denote the universal constant from Lemma 21. Let D denote the maximum of the corresponding universal constants from

Lemmas 15 and 11. Set

$$\begin{aligned} K_2 &:= 1 \vee \frac{1}{(\log 2)^{1/4}} \vee \left(\frac{2}{c}\right)^{1/4} \vee \left(\frac{L^*}{c}\right)^{1/4}, \\ K_3 &:= \frac{L^*}{K_2^2}, \\ K'_2 &:= \frac{1}{\sqrt{\log 2}} \vee \left(\frac{2}{c'}\right)^{1/2} \vee (c' K_3^2)^{-1/4}. \end{aligned}$$

Here, $c := c^* \wedge c^{**}$ with c^* and c^{**} being the universal constants in the exponential terms of Lemmas 15 and 14 respectively. Likewise, $c' := (c')^* \wedge (c')^{**}$ where $(c')^*$ and $(c')^{**}$ are the universal constants in the exponential terms of Lemmas 11 and 10. Note that these choices of K_2, K_3, K'_2 are almost identical to the choices in the proofs of Propositions 2 and 3. The only modifications are the terms $(2/c)^{1/4}$ and $(2/c')^{1/4}$, and the utility of this modification will become clear through the course of the proof.

For any ν , define

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

We examine the Type I and Type II errors separately. Focusing on the Type I error, union bound yields

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

We bound each term separately.

Type I error: Bulk

For $s \in \mathcal{S}_{\text{bulk}}$,

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

Consider that Lemma 15 gives us the following. If we select $x > 0$ satisfying

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

then we have $P_0 \{ \varphi_{\nu,s} = 1 \} \leq e^{-x}$. Let us select

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

To verify (40), consider that $c^* K_2^4 \geq 2$ by our choice of K_2 , and so

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

Here, we have used that $d_\nu = \nu \vee \lceil D \rceil$. We have also used $C \geq 1$, which holds provided we select C_η large enough (i.e. $C_\eta \geq 1$).

Likewise, consider

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

Therefore, (40) is satisfied, and so we have

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

where κ is the universal constant $\kappa = \frac{1}{4(C^* \vee (C^*)^2)} \cdot \frac{1}{(2K_2^4 \lceil D \rceil) \vee (\sqrt{2}/K_2^2)}$. With this bound in hand, observe

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

Here we have used $\log(e|\mathcal{V}_\mathcal{H}|) \leq 2\mathcal{A}_\mathcal{H}$ from Lemma 3. We bound each term separately. First, consider

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

provided we take C_η sufficiently large.

Likewise, consider

$$\begin{aligned}
e^{2\mathcal{A}_{\mathcal{H}}} \sum_{\substack{k \in \mathbb{N} \cup \{0\}: \\ 2^k < \sqrt{p\mathcal{A}_{\mathcal{H}}}}} \exp \left(-C\kappa 2^k \log \left(1 + \frac{p\mathcal{A}_{\mathcal{H}}}{2^{2k}} \right) \right) &\leq e^{2\mathcal{A}_{\mathcal{H}}} \sum_{\substack{k \in \mathbb{N} \cup \{0\}: \\ 2^k < \sqrt{p\mathcal{A}_{\mathcal{H}}}}} \exp \left(-\frac{C\kappa}{2} 2^k \log \left(1 + \frac{p\mathcal{A}_{\mathcal{H}}}{2^{2k}} \right) - \frac{C\kappa}{2} \mathcal{A}_{\mathcal{H}} \right) \\
&\leq e^{-(\frac{C\kappa}{2}-2)\mathcal{A}_{\mathcal{H}}} \sum_{\substack{k \in \mathbb{N} \cup \{0\}: \\ 2^k < \sqrt{p\mathcal{A}_{\mathcal{H}}}}} \exp \left(-\frac{C\kappa}{2} 2^k \log \left(1 + \frac{p\mathcal{A}_{\mathcal{H}}}{2^{2k}} \right) \right) \\
&\leq \frac{\eta}{12}
\end{aligned}$$

again, provided we take C_{η} sufficiently large. We have used that $\mathcal{A}_{\mathcal{H}}$ exhibits at most logarithmic growth in p , i.e. we use the crude bound $\mathcal{A}_{\mathcal{H}} \leq \log(ep)$. Therefore, we have established

$$\sum_{\nu \in \mathcal{V}_{\mathcal{H}}} \sum_{s \in \mathcal{S}_{\text{bulk}}} P_0 \{ \varphi_{\nu,s} = 1 \} \leq \frac{\eta}{6}. \quad (41)$$

Type I error: Tail

For $s \in \mathcal{S}_{\text{tail}}$, we have

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

Consider that Lemma 11 gives us an analogue to (40), namely if we select $x > 0$ satisfying

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

then we have $P_0 \{ \varphi_{\nu,s} = 1 \} \leq e^{-x}$. Let us select

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

To verify (42), consider that $c^{**}(K'_2)^2 \geq 2$ by our choice of K'_2 , and so

$$\begin{aligned}
\sqrt{p(r'_s)^4 e^{-c^{**}(r'_s)^2} x} &= (K'_2)^2 \log \left(1 + \frac{p\mathcal{A}_{\mathcal{H}}}{s^2} \right) \sqrt{p \frac{s^4}{(s^2 + p\mathcal{A}_{\mathcal{H}})^2} x} \\
&\leq \sqrt{C} \frac{1}{2C^{**}} \log \left(1 + \frac{p\mathcal{A}_{\mathcal{H}}}{s^2} \right) \sqrt{p \frac{s^4}{(s^2 + p\mathcal{A}_{\mathcal{H}})^2} \cdot \frac{p\mathcal{A}_{\mathcal{H}}^2}{s^2}} \\
&\leq \frac{C}{2C^{**}} s \log \left(1 + \frac{p\mathcal{A}_{\mathcal{H}}}{s^2} \right).
\end{aligned}$$

Here, we have used $C \geq 1$. Likewise, consider that

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

Therefore, (42) is satisfied, and so we have

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

where κ is the universal constant $\kappa = \frac{1}{4} \cdot \frac{1}{2((C^{**}) \vee (C^{**})^2)}$. With this bound in hand, observe

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

Here we have used $\log(e|\mathcal{V}_{\mathcal{H}}|) \leq 2\mathcal{A}_{\mathcal{H}}$ from Lemma 3. From here, the same argument from the bulk case can be employed to conclude

$$\sum_{\nu \in \mathcal{V}_{\mathcal{H}}} \sum_{s \in \mathcal{S}_{\text{tail}}} P_0 \{ \varphi_{\nu, s} = 1 \} \leq \frac{\eta}{6} \quad (43)$$

provided C_{η} is taken sufficiently large.

Type I error: Dense

Let us now bound $\sum_{\nu \in \mathcal{V}_{\mathcal{H}}} P_0 \{ \varphi_{\nu, p} = 1 \}$. Consider that for any $\nu \in \mathcal{V}_{\mathcal{H}}$, we have

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

To obtain the fourth line, we have used $\mathcal{A}_{\mathcal{H}} \leq \log(ep)$ which implies $\sqrt{\nu p \mathcal{A}_{\mathcal{H}}} \geq \mathcal{A}_{\mathcal{H}}$. In the above display, we have also used $\frac{C}{2} \geq 1$ (which holds provided we take C_{η} large enough) to obtain the penultimate line and Lemma 16 to obtain the final line. Consequently,

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

since $\log(e|\mathcal{V}_{\mathcal{H}}|) \leq 2\mathcal{A}_{\mathcal{H}}$ by Lemma 3 and $\frac{C}{4} - 2 \geq \log(6/\eta)$, which holds provided we take C_{η} sufficiently large.

Putting together (41), (43), (44) into (39) yields the Type I error bound

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

Type II error: We now examine the Type II error. Suppose $s^* \in [p]$. Let $f \in \mathcal{F}_{s^*}$ with $\|f\|_2 \geq C\tau_{\text{adapt}}(p, s^*, n)$. We proceed by considering various cases.

Type II error: Dense

Suppose $s^* \geq \sqrt{p\mathcal{A}_{\mathcal{H}}}$. Let $\tilde{\nu}$ denote the smallest element in $\mathcal{V}_{\mathcal{H}}$ which is greater than or equal to

$\nu_{\mathcal{H}}(s^*, \mathcal{A}_{\mathcal{H}})$. Then

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

Consider that

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

where the collection $\{\theta_{k,j}\}$ denotes the collection of basis coefficients for f . We will also use the matrix Θ to denote the collection of basis coefficients; note that $\Theta \in \mathcal{T}_{s^*}$. Let S^* denote the set of j for which Θ_j are nonzero. Observe

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

Therefore,

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

We have used that $\tilde{\nu} \leq 2\nu_{\mathcal{H}}(s^*, \mathcal{A}_{\mathcal{H}})$ to obtain the third line. Therefore, we have

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

The signal magnitude satisfies the requisite strength needed to successfully detect, which can be seen by following the argument of Proposition 4. Hence, we can achieve Type II error less than or equal to $\frac{\eta}{2}$ by selecting C_{η} sufficiently large. We can combine this bound with (39) to conclude that the total testing risk is bounded above by η , as desired.

Type II error: Bulk

Suppose $s^* < \sqrt{p\mathcal{A}_{\mathcal{H}}}$ and $\sqrt{\log \left(1 + \frac{p\mathcal{A}_{\mathcal{H}}}{(s^*)^2} \right)} \leq K_3 \sqrt{\nu_{\mathcal{H}}(s^*, \mathcal{A}_{\mathcal{H}}) \vee \lceil D \rceil}$. Let $\tilde{\nu}$ be the smallest element in $\mathcal{V}_{\mathcal{H}}$ greater than or equal to $\nu_{\mathcal{H}}(s^*, \mathcal{A}_{\mathcal{H}})$. Let $\tilde{s}$ be the smallest element in $\mathcal{S}$ greater than or equal to s^* . By the definitions of these grids, we have $\tilde{\nu}/2 < \nu_{\mathcal{H}}(s^*, \mathcal{A}_{\mathcal{H}}) \leq \tilde{\nu}$ and $s^* \leq \tilde{s} \leq 2s^*$.

Then

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

With these choices, we have

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

Note that since $\tilde{s} \geq s^*$, we have $f \in \mathcal{F}_{\tilde{s}}$. Following argument similar to those in the proof of Proposition 2, it can be seen that the necessary signal strength to successfully detect is of squared order

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

This is precisely the order of $\tau_{\text{adapt}}^2(p, s^*, n)$, and so we can obtain Type II error less than $\frac{\eta}{2}$ by choosing C_{η} sufficiently large. We can combine this bound with (39) to conclude that the total testing risk is bounded above by η , as desired.

Type II error: Tail

Suppose $s^* < \sqrt{p_{\mathcal{A}_{\mathcal{H}}}}$ and $\sqrt{\log \left(1 + \frac{p_{\mathcal{A}_{\mathcal{H}}}}{(s^*)^2} \right)} > K_3 \sqrt{\nu_{\mathcal{H}}(s^*, \mathcal{A}_{\mathcal{H}}) \vee \lceil D \rceil}$. Let $\tilde{\nu}$ and $\tilde{s}$ be as defined in the previous case, i.e. Type II error analysis for the bulk case. As before, we have

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

Now, consider that $s^* \leq \tilde{s}$ and so $f \in \mathcal{F}_{\tilde{s}}$.

Suppose we have

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

We can follow arguments similar to those in the proof of Proposition 3 to see that the necessary signal strength to successfully detect is of squared order

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

This is precisely the order of $\tau_{\text{adapt}}^2(p, s^*, n)$ and so we can obtain Type II error less than $\frac{\eta}{2}$ by choosing C_{η} sufficiently large.

On the other hand, suppose we have

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

As argued in the previous section about the bulk regime, the necessary signal strength to successfully detect is of squared order

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

Consider that

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

and so we can obtain Type II error less than $\frac{\eta}{2}$ by choosing C_{η} sufficiently large. We can combine this bound with (39) to conclude that the total testing risk is bounded above by η , as desired. $\square$

7.5 Smoothness and sparsity adaptive lower bounds

Proof of Theorem 5. Fix $\eta \in (0, 1)$. Fix any $s \geq p^{1/2+\delta} \sqrt{\log \log n}$. Through the course of the proof, we will note where we must take c_η suitably small enough, so for now let $0 < c < c_\eta$. For each s , define the geometric grid

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

Note $\log |\mathcal{V}_s| \asymp \log \log(np)$.

We now define a prior π which is supported on the alternative hypothesis. Let $\nu \sim \text{Uniform}(\mathcal{V}_s)$. Let $\alpha = \alpha(\nu, s)$ denote the solution to

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

Note that $\alpha(\nu, s) \in [\alpha_0, \alpha_1]$ for any $\nu \in \mathcal{V}_s$. Note α is random since ν is random. Draw uniformly at random a subset $S \subset [p]$ of size s . Draw independently

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

Here, ρ_ν is given by $\rho_\nu := \left(\frac{ns}{\sqrt{p \log \log(np)}} \right)^{-\frac{2\alpha+1}{4\alpha+1}}$. Having defined ρ_ν , the definition of the prior π is complete.

Now we must show π is indeed supported on the alternative hypothesis. Consider that for $\Theta \sim \pi$, we have

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

Furthermore, consider that for $i \in S$ we have by definition of $\alpha = \alpha(\nu, s)$,

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

We can take $c_\eta \leq 1$ to ensure $c^2 \leq 1$. Of course, for $i \in S^c$ we have $\Theta_i = 0$. Hence, we have shown $\Theta \in \mathcal{T}(s, \alpha)$ with probability one, and so π has the proper support.

Writing $P_\pi = \int P_\Theta \pi(d\Theta)$ to denote the mixture induced by the prior π , we have

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

For the following calculations, let $\Theta, \Theta' \stackrel{iid}{\sim} \pi$. Let $\nu, \nu' \in \mathcal{V}_s$ be the corresponding quantities and let α, α' denote the corresponding smoothness levels. Further let S, S' denote the corresponding support sets. Note both are of size s . Let $\{r_{i\ell}, r'_{i\ell}\}_{1 \leq i \leq p, \ell \in \mathbb{N}}$ denote an iid collection of Rademacher(1/2) random variables which is independent of all the other random variables. By the Ingster-Suslina

method (Proposition 5), we have

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

From the definition of $\alpha(\nu, s)$, we have

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

Here, we have used $s \geq \sqrt{p \log \log(np)}$ since $s \geq p^{1/2+\delta}$ and $s \geq \sqrt{p \log \log n}$. We have also used the inequality $x/2 \leq \log(1+x)$ for $0 \leq x \leq 1$. With this in hand, it follows that

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

Noting that we can write $\nu = 2^k$, $\nu' = 2^{k'}$ and that $|\mathcal{V}_s| \asymp \log(np)$, we can follow the same steps as in the proof of Proposition 4.2 in [16]. Taking c_η suitably small, we obtain $\chi^2(P_\pi \| P_0) \leq 4\eta^2$ which yields

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

as desired. $\square$

Proof of Theorem 6. Some simplification is convenient. Note τ_{sparse} can be rewritten as

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

Note that in the case $\log(p \log \log n) > n^{\frac{1}{2\alpha+1}}$, we have $\log p \gtrsim n^{\frac{1}{2\alpha+1}}$. Therefore, $\log(p \log \log n) \asymp \log p$ and so we can further simplify

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

Writing τ_{sparse} in the form (46) is convenient. The lower bound $\frac{s \log p}{n}$ is exactly the minimax lower bound and so no new argument is needed. All that needs to be proved is the lower bound $s \left(n / \sqrt{\log(p \log \log n)} \right)^{-\frac{4\alpha}{4\alpha+1}}$.

The proof is very similar to that of the proof of Theorem 5, so we only point out the modifications in the interest of brevity. Fix any $s < p^{1/2-\delta}$. Let π be the prior from the proof of Theorem 5, but use $\mathcal{V}_{\text{sparse}}$ given by (47) and $\alpha(\nu)$ given by (48), defined below, instead. Define the geometric grid

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

Let $\alpha(\nu)$ denote the solution to

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

Note $\alpha(\nu) \in [\alpha_0, \alpha_1]$ for any $\nu \in \mathcal{V}_{\text{sparse}}$. Also, use

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

It can be checked in the same manner that π with these modifications is properly supported on the alternative hypothesis. We can continue along as in the proof of Theorem 5 up until the point we have

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

From here, we use the fact that $s \leq p^{\frac{1}{2}-\delta}$ implies $\log(p \log \log n) \asymp \log \left(1 + \frac{p \log \log n}{s^2} \right)$. In other words, there exists a constant $\kappa > 0$ such that

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

Then by taking c_η sufficiently small, the rest of the proof of Theorem 5 can be carried out to obtain the desired result. $\square$

7.6 Smoothness and sparsity adaptive upper bounds

Proof of Theorem 4. Fix $\eta \in (0, 1)$. For ease, let us just write $\mathcal{V} = \mathcal{V}_{\text{test}}$. We will note throughout the proof where we take C_η suitably large, so for now let $C > C_\eta$. We will also note where we take K_η suitably large. We will also note where we take K_η suitably large. We first examine the Type I error. By union bound and taking $K_\eta \geq 1$,

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

for some universal positive constant κ . Here, we have used $|\mathcal{V}| \asymp \log(np)$. Taking K_η suitably large, the Type I error is suitably bounded

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

We now examine the Type II error. Fix any $s^* \geq p^{1/2+\delta}$, $\alpha^* \in [\alpha_0, \alpha_1]$, and $f^* \in \mathcal{F}(s^*, \alpha^*)$ with $\|f^*\|_2 \geq C\tau_{\text{dense}}(p, s^*, n, \alpha^*)$. Set

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

Let $\nu \in \mathcal{V}$ be the smallest element larger than ν^* . Note $\nu/2 \leq \nu^* \leq \nu$ by definition of $\mathcal{V}$.

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

Consider that since $\nu \geq \left(\frac{ns^*}{\sqrt{p \log \log(np)}} \right)^{\frac{2}{4\alpha^*+1}}$ and $s^* \geq p^{1/2+\delta}$, it immediately follows that ν grows polynomially in n , and so $\nu \gtrsim \log \log n$. Therefore, we have $\sqrt{\nu p \log \log(np)} \geq \frac{1}{\kappa'} \log \log(np)$ for some universal positive constant κ' . Then by Chebyshev's inequality

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

Consider that

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

Therefore, taking C_{η} sufficiently large, we have

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

Hence, the Type II error is bounded suitably. Since f^* was arbitrary, we have shown that

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

as desired. □

References

- [1] E. Arias-Castro, E. J. Candès, and Y. Plan. “Global testing under sparse alternatives: ANOVA, multiple comparisons and the higher criticism”. *Ann. Statist.* 39.5 (2011), pp. 2533–2556.
- [2] Y. Baraud. “Non-asymptotic minimax rates of testing in signal detection”. *Bernoulli* 8.5 (2002), pp. 577–606.
- [3] S. Bhattacharya, J. Fan, and D. Mukherjee. “Deep neural networks for nonparametric interaction models with diverging dimension”. *Ann. Statist.* (Forthcoming).
- [4] L. Birgé and P. Massart. “Gaussian model selection”. *J. Eur. Math. Soc. (JEMS)* 3.3 (2001), pp. 203–268.
- [5] L. D. Brown and M. G. Low. “Asymptotic equivalence of nonparametric regression and white noise”. *Ann. Statist.* 24.6 (1996), pp. 2384–2398.
- [6] M. V. Burnasev. “Minimax detection of an imperfectly known signal against a background of Gaussian white noise”. *Teor. Veroyatnost. i Primenen.* 24.1 (1979), pp. 106–118.
- [7] A. Carpentier and N. Verzelen. “Adaptive estimation of the sparsity in the Gaussian vector model”. *Ann. Statist.* 47.1 (2019), pp. 93–126.
- [8] J. Chhor, R. Mukherjee, and S. Sen. “Sparse signal detection in heteroscedastic Gaussian sequence models: sharp minimax rates”. *Bernoulli* 30.3 (2024), pp. 2127–2153.
- [9] O. Collier, L. Comminges, and A. B. Tsybakov. “On estimation of nonsmooth functionals of sparse normal means”. *Bernoulli* 26.3 (2020), pp. 1989–2020.
- [10] O. Collier, L. Comminges, and A. B. Tsybakov. “Minimax estimation of linear and quadratic functionals on sparsity classes”. *Ann. Statist.* 45.3 (2017), pp. 923–958.
- [11] O. Collier, L. Comminges, A. B. Tsybakov, and N. Verzelen. “Optimal adaptive estimation of linear functionals under sparsity”. *Ann. Statist.* 46.6A (2018), pp. 3130–3150.
- [12] L. Comminges, O. Collier, M. Ndaoud, and A. B. Tsybakov. “Adaptive robust estimation in sparse vector model”. *Ann. Statist.* 49.3 (2021), pp. 1347–1377.
- [13] N. Deb, R. Mukherjee, S. Mukherjee, and M. Yuan. “Detecting structured signals in Ising models”. *Ann. Appl. Probab.* 34.1A (2024), pp. 1–45.
- [14] D. Donoho and J. Jin. “Higher criticism for detecting sparse heterogeneous mixtures”. *Ann. Statist.* 32.3 (2004), pp. 962–994.
- [15] M. S. Ermakov. “Minimax detection of a signal in Gaussian white noise”. *Teor. Veroyatnost. i Primenen.* 35.4 (1990), pp. 704–715.
- [16] C. Gao, F. Han, and C.-H. Zhang. “On estimation of isotonic piecewise constant signals”. *Ann. Statist.* 48.2 (2020), pp. 629–654.

- [17] G. Gayraud and Y. Ingster. “Detection of sparse additive functions”. *Electron. J. Stat.* 6 (2012), pp. 1409–1448.
- [18] P. Hall and J. Jin. “Innovated higher criticism for detecting sparse signals in correlated noise”. *Ann. Statist.* 38.3 (2010), pp. 1686–1732.
- [19] Y. Han, J. Jiao, and R. Mukherjee. “On estimation of L_r -norms in Gaussian white noise models”. *Probab. Theory Related Fields* 177.3-4 (2020), pp. 1243–1294.
- [20] T. Hastie and R. Tibshirani. “Generalized additive models”. *Statist. Sci.* 1.3 (1986), pp. 297–318.
- [21] Y. I. Ingster. “Adaptive tests for minimax testing of nonparametric hypotheses”. *Probability theory and mathematical statistics, Vol. I (Vilnius, 1989)*. “Mokslas”, Vilnius, 1990, pp. 539–549.
- [22] Y. I. Ingster. “Asymptotically minimax testing of nonparametric hypotheses”. *Probability theory and mathematical statistics, Vol. I (Vilnius, 1985)*. VNU Sci. Press, Utrecht, 1987, pp. 553–574.
- [23] Y. I. Ingster. “Minimax nonparametric detection of signals in white Gaussian noise”. *Problemy Peredachi Informatsii* 18.2 (1982), pp. 61–73.
- [24] Y. I. Ingster and I. A. Suslina. *Nonparametric Goodness-of-Fit Testing Under Gaussian Models*. Vol. 169. Lecture Notes in Statistics. Springer-Verlag, New York, 2003.
- [25] Y. Ingster and O. Lepski. “Multichannel nonparametric signal detection”. *Math. Methods Statist.* 12.3 (2003), pp. 247–275.
- [26] Y. I. Ingster, A. B. Tsybakov, and N. Verzelen. “Detection boundary in sparse regression”. *Electron. J. Stat.* 4 (2010), pp. 1476–1526.
- [27] I. M. Johnstone. “Gaussian estimation: Sequence and wavelet models”.
- [28] I. M. Johnstone. “Chi-square oracle inequalities”. *State of the art in probability and statistics (Leiden, 1999)*. Vol. 36. IMS Lecture Notes Monogr. Ser. Inst. Math. Statist., Beachwood, OH, 2001, pp. 399–418.
- [29] V. Koltchinskii and M. Yuan. “Sparsity in multiple kernel learning”. *Ann. Statist.* 38.6 (2010), pp. 3660–3695.
- [30] S. Kotekal and C. Gao. “Minimax rates for sparse signal detection under correlation”. *Inf. Inference* 12.4 (2023), Paper No. iaad044, 97.
- [31] B. Laurent and P. Massart. “Adaptive estimation of a quadratic functional by model selection”. *Ann. Statist.* 28.5 (2000), pp. 1302–1338.
- [32] O. Lepski, A. Nemirovski, and V. Spokoiny. “On estimation of the L_r norm of a regression function”. *Probab. Theory Related Fields* 113.2 (1999), pp. 221–253.
- [33] Y. Lin and H. H. Zhang. “Component selection and smoothing in multivariate nonparametric regression”. *Ann. Statist.* 34.5 (2006), pp. 2272–2297.
- [34] H. Liu, C. Gao, and R. J. Samworth. “Minimax rates in sparse, high-dimensional change point detection”. *Ann. Statist.* 49.2 (2021), pp. 1081–1112.
- [35] L. Meier, S. van de Geer, and P. Bühlmann. “High-dimensional additive modeling”. *Ann. Statist.* 37.6B (2009), pp. 3779–3821.
- [36] R. Mukherjee and G. Ray. “On testing for parameters in Ising models”. *Ann. Inst. Henri Poincaré Probab. Stat.* 58.1 (2022), pp. 164–187.

- [37] R. Mukherjee and S. Sen. “On minimax exponents of sparse testing”. arXiv:2003.00570 [math, stat]. 2020.
- [38] G. Raskutti, M. J. Wainwright, and B. Yu. “Minimax-optimal rates for sparse additive models over kernel classes via convex programming”. *J. Mach. Learn. Res.* 13 (2012), pp. 389–427.
- [39] P. Ravikumar, J. Lafferty, H. Liu, and L. Wasserman. “Sparse additive models”. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 71.5 (2009), pp. 1009–1030.
- [40] M. Reiß. “Asymptotic equivalence for nonparametric regression with multivariate and random design”. *Ann. Statist.* 36.4 (2008), pp. 1957–1982.
- [41] V. G. Spokoiny. “Adaptive hypothesis testing using wavelets”. *Ann. Statist.* 24.6 (1996), pp. 2477–2498.
- [42] N. M. Temme. “The asymptotic expansion of the incomplete gamma functions”. *SIAM J. Math. Anal.* 10.4 (1979), pp. 757–766.
- [43] N. M. Temme. “Uniform asymptotic expansions of the incomplete gamma functions and the incomplete beta function”. *Math. Comp.* 29.132 (1975), pp. 1109–1114.
- [44] A. B. Tsybakov. “Pointwise and sup-norm sharp adaptive estimation of functions on the Sobolev classes”. *Ann. Statist.* 26.6 (1998), pp. 2420–2469.
- [45] A. B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer Series in Statistics. Springer, New York, 2009.
- [46] R. Vershynin. *High-Dimensional Probability*. Vol. 47. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2018.
- [47] M. J. Wainwright. *High-Dimensional Statistics*. Vol. 48. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2019.
- [48] Y. Wei and M. J. Wainwright. “The local geometry of testing in ellipses: tight control via localized Kolmogorov widths”. *IEEE Trans. Inform. Theory* 66.8 (2020), pp. 5110–5129.
- [49] Y. Yang and S. T. Tokdar. “Minimax-optimal nonparametric regression in high dimensions”. *Ann. Statist.* 43.2 (2015), pp. 652–674.
- [50] M. Yuan and D.-X. Zhou. “Minimax optimal rates of estimation in high dimensional additive models”. *Ann. Statist.* 44.6 (2016), pp. 2564–2593.
- [51] A. R. Zhang and Y. Zhou. “On the non-asymptotic and sharp lower tail bounds of random variables”. *Stat* 9 (2020), e314, 11.

A Hard thresholding

In this section, we collect results about the random variable $(\|Z\|^2 - \alpha_t(d))\mathbb{1}_{\{\|Z\|^2 \geq d+t^2\}}$ for $Z \sim N(\theta, I_d)$ and $\alpha_t(d)$ defined in (13). We consider the tail $t^2 \gtrsim d$ and the bulk $t^2 \lesssim d$ separately.

The proof outlines are similar to those employed in [34]. However, they only consider $d = 1$, whereas we need to consider the general d case. Consequently, a much more careful analysis is required.

A.1 Tail

Lemma 8 (Moment generating function). *Suppose $\tilde{c}$ is a positive universal constant. There exist universal positive constants $D, \tilde{C}, C$, and c such that if $d \geq D, t^2 \geq \tilde{c}d$, and $0 < \lambda < \frac{1}{2C}$, then we have*

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

where $Y = (\|Z\|^2 - \alpha_t(d)) \mathbb{1}_{\{\|Z\|^2 \geq d+t^2\}}$, $Z \sim N(0, I_d)$, and where $\alpha_t(d)$ is given by (13).

Proof. We follow the broad approach of the argument presented in the proof of Lemma 18 in [34]. The universal positive constant D will be chosen later on in the proof, so for now let $d \geq D$ and $Z \sim N(0, I_d)$. Note we will also take D large enough so that Lemma 21 is applicable. Since $E(Y) = 0$, we have

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

Consider that for any $y \in \mathbb{R}$ we have

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

Therefore,

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

Each term will be bounded separately. Considering the first term, note that Lemma 22 asserts $\alpha_t(d) \leq d + C^*(t^2 \vee d)$ for a universal constant C^* . Note also that $\alpha_t(d) \geq d + t^2$. Therefore,

$$\begin{aligned} \lambda^2 E(Y^2 \mathbb{1}_{\{Y < 0\}}) &= \lambda^2 E((\|Z\|^2 - \alpha_t(d))^2 \mathbb{1}_{\{\alpha_t(d) > \|Z\|^2 \geq d+t^2\}}) \\ &\leq \lambda^2 (\alpha_t(d) - d - t^2)^2 P\{\alpha_t(d) > \|Z\|^2 \geq d+t^2\} \\ &\leq \lambda^2 (d + C^*(t^2 \vee d) - d - t^2)^2 P\{\|Z\|^2 \geq d+t^2\} \\ &\leq 2C^* \lambda^2 (d^2 \vee t^4) \exp\left(-c \min\left(\frac{t^4}{d}, t^2\right)\right) \\ &= 2C^* \lambda^2 t^4 \exp(-ct^2) \end{aligned}$$

where c is a universal positive constant and C^* remains a universal constant but whose value may change from line to line. Note we have also used Corollary 1 in the above display as well as $t^2 \geq \tilde{c}d$. To summarize, we have shown

$$\lambda^2 E(Y^2 \mathbb{1}_{\{Y < 0\}}) \leq 2C^* \lambda^2 t^4 \exp(-ct^2). \quad (50)$$

We now bound the second term in (49). Let f_d denote the probability density function of the χ_d^2 distribution. We have

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

An application of Lemma 18 yields

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

where c remains a universal constant but has a value which may change from line to line. Note we have used $\alpha_t(d) \geq d + t^2$, Corollary 1, and $t^2 \geq \tilde{c}d$. Likewise, consider that

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

where we have used $(a+b)^2 \leq 2a^2 + 2b^2$, $t^2 \geq \tilde{c}d$, Lemma 22, and Corollary 1. Again, C^* and c remain universal constants but have values which may change from line to line. Thus, we have shown

$$\lambda^2 E(Y^2 \mathbb{1}_{\{Y < 0\}}) \leq C^* \lambda^2 t^4 \exp(-ct^2). \quad (51)$$

We now bound the final term in (49). Consider that for $0 \leq \lambda < \frac{1}{2}$ we have

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

Let $u = (\frac{1}{2} - \lambda)z$. Then $du = (\frac{1}{2} - \lambda)dz$ and so

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

where $\Gamma(s, x) = \int_x^{\infty} e^{-t} t^{s-1} dt$. To summarize, we have shown

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

To continue, we would like to apply Corollary 2, but we must first verify the conditions. Let $a = \frac{d}{2}$ and $\eta = \sqrt{2(\mu - \log(1 + \mu))}$ with $\mu = \frac{(\frac{1}{\lambda} + \alpha_t(d))(\frac{1}{2} - \lambda)}{a} - 1$. Note we can take D to be a sufficiently large universal constant in order to ensure a is sufficiently large. Since $\tilde{c}d \leq t^2$, it follows from Lemma 22 that $\alpha_t(d) \leq \tilde{C}t^2$ for some positive universal constant $\tilde{C}$. Without loss of generality, we can take

$\tilde{C} \geq \frac{3}{2}$. Let us restrict our attention to $\lambda < \frac{1}{2\tilde{C}}$. Observe that

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

where we have defined $C^{**} = 2(\tilde{C} - 1)$. Since we have shown $\mu > 0$, we can apply Corollary 2. By Corollary 2 we have

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

Here, Φ denotes the cumulative distribution function of the standard normal distribution and C^* is a universal constant. By the Gaussian tail bound $1 - \Phi(x) \leq \frac{\varphi(x)}{x}$ for $x > 0$ where $\varphi = \Phi'$, we have

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

where the value of C^* has changed from line to line but remains a universal constant. To continue the calculation, we need to bound $\eta\sqrt{a}$ from below. Note we can take $D \geq C^{**}$. For $\lambda < \frac{1}{2\tilde{C}}$, it follows from $\frac{C^{**}}{d} \leq 1$ and that

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

Consequently, we have the bound

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

where the value of C^* has changed but remains a universal constant. We now examine the term $e^{-a\eta^2/2}$. Consider that

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

Therefore, letting the value of C^* change from line to line but remaining a universal constant, we

have from (52)

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

for a universal positive constant c^{**} . We have used that $u - \log(1+u) \gtrsim u$ for $u \geq 1$. Note that we can use this since $\frac{\frac{1}{\lambda} + \alpha_t(d) - d}{d} \geq \frac{t^2}{d} \gtrsim 1$. Since $e^{-\frac{1}{2c^{**}u}} \lesssim u^2$ for all $u \in \mathbb{R}$, it follows that

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

where the value of C^* , again, has changed but remains a universal constant. Putting together our bounds (50), (51), (53) into (49) yields

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

for $\lambda < \frac{1}{2C}$ where $C, c > 0$ are universal constants. The proof is complete. $\square$

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

$$E\left\{(|Z|^2 - \alpha_t(d)) \mathbb{1}_{\{|Z|^2 \geq d+t^2\}}\right\} \begin{cases} = 0 & \text{if } \theta = 0, \\ \geq 0 & \text{if } \|\theta\|^2 < Ct^2, \\ \geq \|\theta\|^2/2 & \text{if } \|\theta\|^2 \geq Ct^2. \end{cases}$$

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

Proof. We will make a choice for C at the end of the proof. Consider first the case where $\theta = 0$. Then $E\left\{(|Z|^2 - \alpha_t(d)) \mathbb{1}_{\{|Z|^2 \geq d+t^2\}}\right\} = 0$ by definition of $\alpha_t(d)$ and so we have the desired result. The second case follows since the expression inside the expectation is stochastically increasing in $\|\theta\|$. Moving on to the final case, suppose $\|\theta\|^2 \geq Ct^2$. Since $\alpha_t(d) \geq d + t^2$, we have

$$E\left\{(|Z|^2 - \alpha_t(d)) \mathbb{1}_{\{|Z|^2 < d+t^2\}}\right\} \leq 0$$

Consequently,

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

By Lemma 22 and $t^2 \gtrsim d$, there exists a universal positive constant C^* such that $\alpha_t(d) \leq d + C^*t^2$. Therefore,

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

Selecting $C = \frac{C^*}{2}$ yields the desired result. $\square$

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

$$\text{Var}((\|Z\|^2 - \alpha_t(d))\mathbb{1}_{\{\|Z\|^2 \geq d+t^2\}}) \lesssim \begin{cases} t^4 \exp\left(-c^* \min\left(\frac{t^4}{d}, t^2\right)\right) & \text{if } \theta = 0, \\ t^4 & \text{if } 0 < \|\theta\| < 2t, \\ \|\theta\|^2 & \text{if } \|\theta\| > 2t. \end{cases}$$

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

Proof. The proof of Lemma 14 can be repeated with the modification of invoking Lemma 22 instead of Lemma 21. $\square$

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

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

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

Proof. Let D be the constant from Lemma 8. We use the Chernoff method to obtain the desired bound. Let Y be as in Lemma 8 and let $\tilde{C}$ be the universal constant from Lemma 8. For any $u > 0$, we have by Lemma 8

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

where $c > 0$ is a universal constant and $C > 0$ is a universal constant whose value may change from line to line but remain a universal constant. Selecting $u = \frac{1}{2} \left(\sqrt{\frac{xnt^4 e^{-ct^2}}{C}} + \frac{x}{C}\right)$ and selecting C^* suitably yields the desired result. The proof is complete. $\square$

A.2 Bulk

Lemma 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, $Y = (\|Z\|^2 - \alpha_t(d))\mathbb{1}_{\{\|Z\|^2 \geq d+t^2\}}$ with $t^2 = \beta\sqrt{d}$ and $Z \sim N(0, I_d)$. Also, $\alpha_t(d)$ is given by (13).

Proof. The argument we give will be similar to the proof of Lemma 8, namely we will separately bound each term in (49) and substitute into the equation

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

We start with the first term on the right-hand side in (49). By Lemma 18, we have $\alpha_t(d) \leq d + C^* \beta\sqrt{d}$ for a universal positive constant C^* . Note that $C^* \geq 1$ since we trivially have $\alpha_t(d) \geq d + t^2 = d + \beta\sqrt{d}$.

Consequently, arguing as in the proof of Lemma 8, we have

$$\begin{aligned}\lambda^2 E(Y^2 \mathbb{1}_{\{Y < 0\}}) &\leq \lambda^2 (d + t - \alpha_t(d))^2 P\{\|Z\|^2 \geq d + t^2\} \\ &\leq C\lambda^2 d\beta^2 \exp\left(-c \min\left(\frac{t^4}{d}, t^2\right)\right) \\ &= C\lambda^2 d\beta^2 e^{-c\beta^2}\end{aligned}$$

where $C, c > 0$ are universal constants. We have used Corollary 1 to obtain the final inequality. We now bound the second term in (49). Letting f_d denote the probability density function of the χ_d^2 distribution, we can repeat and then continue the calculation in the proof of Lemma 8 to obtain

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

where the values of $C, c > 0$ have changed but remain universal constants. We have used Lemma 24 here.

We now bound the final term in (49). Arguing exactly as in the proof of Lemma 8, we have

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

for $0 \leq \lambda < \frac{1}{2}$. Recall that $\Gamma(s, x) = \int_x^{\infty} e^{-t} t^{s-1} dt$.

Let us restrict our attention to $\lambda < \frac{\beta}{2(\sqrt{d} + C^* \beta)}$. We seek to apply Corollary 2, but we must verify the conditions. Let $a = \frac{d}{2}$ and $\eta = \sqrt{2(\mu - \log(1 + \mu))}$ with $\mu = \frac{(\frac{1}{\lambda} + \alpha_t(d))(\frac{1}{2} - \lambda)}{a} - 1$. Observe

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

Consider that $\lambda < \frac{\beta}{2(\sqrt{d} + C^* \beta)} = \frac{t^2}{2(d + C^* \beta \sqrt{d})} \leq \frac{\alpha_t(d) - d}{2\alpha_t(d)}$. Therefore, $\alpha_t(d)(1 - 2\lambda) - d \geq 0$ and so

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

Note we have used $C^* \geq 1$. Since $\mu > 0$ and $d \geq C^{**}$ which is a sufficiently large universal constant, we can apply Corollary 2. By Corollary 2, we have

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

Here, $C^\dagger$ is a positive universal constant. Recall that Φ denotes the cumulative distribution function of the standard normal distribution. By the Gaussian tail bound $1 - \Phi(x) \leq \frac{\varphi(x)}{x}$ for $x > 0$, we have

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

where the value of $C^\dagger$ has changed but remains a universal constant. To continue with the bound,

we need to bound $\eta\sqrt{a}$ from below. Let us take C^{**} larger than 4. Since $a = \frac{d}{2}$, we have

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

where $c^* > 0$ is a universal constant. We can conclude from (54) that

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

where the value of $C^\dagger$ has changed but remains a universal constant. We now examine the term $e^{-\frac{a\eta^2}{2}}$. Arguing exactly as in the proof of Lemma 8, we obtain

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

which, as argued in the proof of Lemma 8, leads us to the bound

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

for $\lambda < \frac{\beta}{2(\sqrt{d} + C^*\beta)}$ as desired. $\square$

Lemma 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*

$$E\left\{(|Z|^2 - \alpha_t(d))\mathbb{1}_{\{|Z|^2 \geq d+t^2\}}\right\} \begin{cases} = 0 & \text{if } \theta = 0, \\ \geq 0 & \text{if } \|\theta\|^2 < Ct^2, \\ \geq \|\theta\|^2/2 & \text{if } \|\theta\|^2 \geq Ct^2 \end{cases}$$

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

Proof. We will make a choice for C later on in the proof. The proof for the cases $\theta = 0$ and $\|\theta\|^2 \leq Ct^2$ follows exactly as in the proof of Lemma 9. Here, we focus on the final case in which $\|\theta\|^2 \geq Ct^2$. Since $\alpha_t(d) \geq d + t^2$, we have

$$E\left\{(|Z|^2 - \alpha_t(d))\mathbb{1}_{\{|Z|^2 < d+t^2\}}\right\} \leq 0.$$

Since $t^2 = \beta\sqrt{d}$, we can apply Lemma 21 to obtain

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

where C^* is a universal positive constant. Taking $C = \frac{C^*}{2}$ completes the proof. $\square$

Lemma 14. Let $Z \sim N(\theta, I_d)$. Suppose $1 \leq \beta \leq L^* \sqrt{d}$ 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

$$\text{Var}((\|Z\|^2 - \alpha_t(d)) \mathbb{1}_{\{\|Z\|^2 \geq d+t^2\}}) \lesssim \begin{cases} t^4 \exp\left(-\frac{c^* t^4}{d}\right) & \text{if } \theta = 0, \\ d + \|\theta\|^2 & \text{if } \|\theta\| \geq 2t, \\ t^4 & \text{if } 0 < \|\theta\| < 2t. \end{cases}$$

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

Proof. Let f_d and F_d respectively denote the probability density function and cumulative distribution function of the χ_d^2 distribution.

Case 1: Consider the first case in which $\theta = 0$. Then

$$\begin{aligned} \text{Var}((\|Z\|^2 - \alpha_t(d)) \mathbb{1}_{\{\|Z\|^2 \geq d+t^2\}}) &\leq E((\|Z\|^2 - \alpha_t(d))^2 \mathbb{1}_{\{\|Z\|^2 \geq d+t^2\}}) \\ &= P\{\|Z\|^2 \geq d+t^2\} \cdot E((\|Z\|^2 - \alpha_t(d))^2 \mid \|Z\|^2 \geq d+t^2) \\ &= (1 - F_d(d+t^2)) \text{Var}(\|Z\|^2 \mid \|Z\|^2 \geq d+t^2) \\ &\leq C^* (1 - F_d(d+t^2)) d \beta^2 \end{aligned}$$

where we have applied the definition of $\alpha_t(d)$ and we have applied Lemma 24. Here, C^* is a universal positive constant. An application of Corollary 1 gives the desired result for this case.

We now move to the other two cases. Suppose $\theta \neq 0$. For ease of notation, let $Y = (\|Z\|^2 - \alpha_t(d)) \mathbb{1}_{\{\|Z\|^2 \geq d+t^2\}}$. Observe

$$\begin{aligned} \text{Var}(Y) &= E(\text{Var}(Y \mid \mathbb{1}_{\{\|Z\|^2 \geq d+t^2\}})) + \text{Var}(E(Y \mid \mathbb{1}_{\{\|Z\|^2 \geq d+t^2\}})) \\ &\leq P\{\|Z\|^2 \geq d+t^2\} \text{Var}(\|Z\|^2 \mid \|Z\|^2 \geq d+t^2) + P\{\|Z\|^2 \geq d+t^2\} (E(\|Z\|^2 - \alpha_t(d) \mid \|Z\|^2 \geq d+t^2))^2. \end{aligned} \quad (55)$$

Note that $0 \leq E(\|Z\|^2 - \alpha_t(d) \mid \|Z\|^2 \geq d+t^2)$ since $\theta \neq 0$ (by an appeal to stochastic ordering), and so we can find upper bounds for the square of this conditional expectation by first finding upper bounds on the conditional expectation. We examine each term separately in the above display. First, consider

$$\begin{aligned} P\{\|Z\|^2 \geq d+t^2\} \text{Var}(\|Z\|^2 \mid \|Z\|^2 \geq d+t^2) &\leq E(\text{Var}(\|Z\|^2 \mid \mathbb{1}_{\{\|Z\|^2 \geq d+t^2\}})) \\ &\leq \text{Var}(\|Z\|^2) \\ &= 2d + 4\|\theta\|^2. \end{aligned} \quad (56)$$

We now examine the second term. Note we can write $Z = g + \theta$ where $g \sim N(0, I_d)$. Therefore,

$$\begin{aligned} E(\|Z\|^2 - \alpha_t(d) \mid \|Z\|^2 \geq d+t^2) &= E(\|g + \theta\|^2 - \alpha_t(d) \mid \|g + \theta\|^2 \geq d+t^2) \\ &\leq E(\|g\|^2 + 2\langle \theta, g \rangle + \|\theta\|^2 - d - t^2 \mid \|g + \theta\|^2 \geq d+t^2) \\ &= E(\|g\|^2 - d \mid \|g + \theta\|^2 \geq d+t^2) + \|\theta\|^2 - t^2 \end{aligned} \quad (57)$$

where we have used that $\langle \theta, g \rangle \mathbb{1}_{\{\|g+\theta\|^2 \geq d+t^2\}} \stackrel{d}{=} \langle \theta, -g \rangle \mathbb{1}_{\{\| -g+\theta \|^2 \geq d+t^2\}}$. With this in hand, we have $E(\langle \theta, g \rangle \mathbb{1}_{\{\|g+\theta\|^2 \geq d+t^2\}}) = 0$, which further implies $E(\langle \theta, g \rangle \mid \|g + \theta\|^2 \geq d+t^2) = 0$. Note we have also used $\alpha_t(d) \geq d+t^2$ to obtain the second line in the above display. With the above display in hand, we now examine the remaining two cases.

Case 2: Consider the case $\|\theta\| \geq 2t$. Observe that

$$E(\|g\|^2 - d \mid \|g + \theta\|^2 \geq d+t^2) = \frac{E((\|g\|^2 - d) \mathbb{1}_{\{\|g+\theta\|^2 \geq d+t^2\}})}{P\{\|g + \theta\|^2 \geq d+t^2\}}.$$

Examining the denominator, since $\|g + \theta\|^2 \sim \chi_d^2(\|\theta\|^2) \stackrel{d}{=} \chi_{d-1}^2 + \chi_1^2(\|\theta\|^2)$ where the two χ^2 -variables on the right hand side are independent, we have

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

where c is a universal positive constant. Examining the numerator, consider that by Lemma 18 we have

$$\begin{aligned} E((\|g\|^2 - d)\mathbb{1}_{\{\|g + \theta\|^2 \geq d + t^2\}}) &\leq E((\|g\|^2 - d)\mathbb{1}_{\{\|g\|^2 \geq d\}}) \\ &= \int_d^\infty (z - d)f_d(z) dz \\ &= d \int_d^\infty (f_{d+2}(z) - f_d(z)) dz \\ &= -2d \int_d^\infty f'_{d+2}(z) dz \\ &= 2df_{d+2}(d). \end{aligned}$$

Hence, from (57), we have shown

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

Thus, we have the bound

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

Consider that

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

where the value of c can change in each expression but remains a universal positive constant. The final inequality follows from Stirling's formula. Moreover, since $\|\theta\| \geq 2t$, there exists a positive universal constant c' such that $P\{\chi_1^2(\|\theta\|^2) \geq 1 + t^2\} \geq 1/c'$. Furthermore, it follows by Chebyshev's inequality that

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

Consequently,

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

The proof for this case is complete.

Case 3: Consider the case $0 < \|\theta\| < 2t$. Then

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

where we have used Lemma 21. Therefore, using the above bound with (56) and plugging into (55), we obtain

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

The proof is complete. $\square$

Lemma 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*

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

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

Proof. The proof of Lemma 11 can be repeated with the modification of invoking Lemma 12 instead of Lemma 8 and taking infimum over $\lambda < \frac{\beta}{2(\beta C^* + \sqrt{d})}$. $\square$

B Properties of the χ_d^2 distribution

Theorem 7 (Bernstein's inequality - Theorem 2.8.1 [46]). *Let $Y_1, \dots, Y_d$ be independent mean-zero subexponential random variables. Then, for every $u \geq 0$, we have*

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

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])

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

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

where $c > 0$ is a universal constant.

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

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

Lemma 17 (Corollary 3 [51]). *Suppose $Z \sim N(0, I_d)$. There exist universal constants $C, c > 0$ such that*

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

for all $u > 0$.

Lemma 18 ([28]). *Let f_d and F_d respectively denote the probability density and cumulative distribution functions of the χ_d^2 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) $(1 - F_{d+2}(t)) - (1 - F_d(t)) = \frac{e^{-t/2} t^{d/2}}{2^{d/2} \Gamma(d/2+1)}$.

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

Proof. By Mean Value Theorem, we have for any $r \geq t$,

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

Consider that $\lim_{r \rightarrow \infty} 1 - F_d(r) = \lim_{r \rightarrow \infty} f_d(r) = 0$. So taking $r \rightarrow \infty$ yields

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

We now evaluate the infimum on the left-hand side. For $x \geq d$, consider that an application of Lemma 18 gives

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

Since $x \geq d > d - 2$, it follows that

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

Thus we can immediately conclude $2f_d(t) \leq 1 - F_d(t)$ as desired. $\square$

Lemma 20. Let F_d denote the cumulative distribution function of the χ_d^2 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. The result follows directly from a change of variables when integrating the probability density function. $\square$

Lemma 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. For ease of notation, let $r^2 = d + t^2$. Let f_d and F_d denote the probability density and cumulative distribution functions of the χ_d^2 distribution. We will select the universal constant L^* later on in the proof. By Lemma 18, we have

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

Rearranging terms and invoking Stirling's approximation (which states $\Gamma(x+1) \sim \sqrt{2\pi x} \left(\frac{x}{e}\right)^x$ as $x \rightarrow \infty$) yields

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

for a universal constant $c > 0$ since d is larger than a sufficiently large universal constant. Applying Lemma 20 and Corollary 2 yields

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

where c^{**} is a universal constant. Observe that $\eta\sqrt{a}$ is larger than a sufficiently large universal constant since d is larger than a sufficiently large universal constant. Using the fact that $1 - \Phi(x) = \frac{1}{x\sqrt{2\pi}} e^{-x^2/2} (1 + o(1))$ as $x \rightarrow \infty$, we have

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

for a universal positive constant c^* . Consider that

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

Consequently,

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

Therefore, we have the bound

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

Consider further that the inequality $x - \log(1 + x) \leq \frac{x^2}{2} \leq x^2$ gives us

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

Hence, we have

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

Let us take $L^* := \frac{c^*}{2c^{**}}$. With this choice, we have $\beta \leq \sqrt{d} \cdot \frac{c^*}{2c^{**}}$ and so $\left(\frac{c^*}{\beta} - \frac{c^{**}}{\sqrt{d}}\right) \geq \frac{c^*}{2\beta}$. Therefore,

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

for a universal positive constant C^* . Thus we have

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

as desired. $\square$

Theorem 8 (Uniform expansion of the incomplete gamma function [43]). *For $a > 0$ and $x \geq 0$ real numbers, 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 $N \geq 0$, we have

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

where the remainder term satisfies

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

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

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

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

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

where

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

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

$$|Q(a, x) - (1 - \Phi(\eta\sqrt{a}))| \leq \frac{C e^{-\frac{a\eta^2}{2}}}{\sqrt{2\pi a}}$$

where C is some universal positive constant.

Proof. The first two displays follow exactly from Theorems 8 and 9. To show the final display, we must show that $\left|\frac{1}{\mu} - \frac{1}{\eta}\right| \lesssim 1$ whenever $\mu > 0$. First, consider the Taylor expansion

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

where ξ is some point between 0 and μ . Therefore,

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

Thus

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

since ξ is between 0 and μ . Therefore, the final display in the statement of the Corollary follows by taking a to be larger than some universal constant and taking C to be a large enough universal constant. $\square$

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

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

where C^* is a positive universal constant.

Proof. We will choose a universal constant $L \geq 1$ at the end of the proof, so for now we leave it as undetermined. Let f_d denote the probability density function of the χ_d^2 distribution. Observe that

$$\begin{aligned} E(\|Z\|^2 \mid \|Z\|^2 \geq d + t^2) &= E(\|Z\|^2 \mathbb{1}_{\{\|Z\|^2 \leq d + Lt^2\}} \mid \|Z\|^2 \geq d + t^2) + E(\|Z\|^2 \mathbb{1}_{\{\|Z\|^2 > d + Lt^2\}} \mid \|Z\|^2 \geq d + t^2) \\ &\leq d + Lt^2 + \sqrt{E(\|Z\|^4 \mid \|Z\|^2 \geq d + t^2)} \sqrt{P\{\|Z\|^2 > d + Lt^2\}}. \end{aligned}$$

By Corollary 1, there exists a universal constant c_1 such that

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

Here we have used $L \geq 1$. Further consider that

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

Therefore, by Lemma 17 we have

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

where C and c_2 are universal positive constants. Taking $L := \frac{c_2}{c_1} \vee 1$ completes the proof. $\square$

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

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

where C^* is a positive universal constant.

Proof. We will choose a universal constant $L \geq 1$ at the end of the proof, so for now we leave it as undetermined. Let f_d denote the probability density function of the χ_d^2 distribution. Consider

$$\begin{aligned} E(\|Z\|^4 \mid \|Z\|^2 \geq d + t^2) &= E(\|Z\|^4 \mathbb{1}_{\{\|Z\|^2 \leq d + Lt^2\}} \mid \|Z\|^2 \geq d + t^2) + E(\|Z\|^4 \mathbb{1}_{\{\|Z\|^2 > d + Lt^2\}} \mid \|Z\|^2 \geq d + t^2) \\ &\leq d^2 + 2Lt^2d + L^2t^4 + \sqrt{E(\|Z\|^8 \mid \|Z\|^2 \geq d + t^2)} \sqrt{P\{\|Z\|^2 > d + Lt^2\}}. \end{aligned}$$

By Corollary 1, there exists a universal positive constant c_1 such that

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

Here we have used $L \geq 1$. Further consider

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

Hence, by Lemma 17 we have

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

where $C > 0$ is a universal constant. Taking $L > \frac{c_2}{c_1} \vee 1$ completes the proof. $\square$

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

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

for some universal positive constant C .

Proof. We will use a universal constant $L \geq 1$ in our proof; a choice for it will be made at the end. Let f_d denote the probability density of the χ_d^2 distribution. Observe that

$$\begin{aligned} &\text{Var}(\|Z\|^2 \mid \|Z\|^2 \geq d + t^2) \\ &= \text{Var}(\|Z\|^2 - d \mid \|Z\|^2 \geq d + t^2) \\ &\leq E((\|Z\|^2 - d)^2 \mid \|Z\|^2 \geq d + t^2) \\ &= E((\|Z\|^2 - d)^2 \mathbb{1}_{\{\|Z\|^2 \leq d + Lt^2\}} \mid \|Z\|^2 \geq d + t^2) + E((\|Z\|^2 - d)^2 \mathbb{1}_{\{\|Z\|^2 > d + Lt^2\}} \mid \|Z\|^2 \geq d + t^2) \\ &\leq L^2t^4 + \sqrt{E((\|Z\|^2 - d)^4 \mid \|Z\|^2 \geq d + t^2)} \sqrt{P\{\|Z\|^2 \geq d + Lt^2\}}. \end{aligned}$$

By Corollary 1, there exists a universal positive constant c_1 such that

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

Here, we have used $L \geq 1$. With this term under control, now observe

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

Hence, by Lemma 17 we have

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

Taking $L > \frac{c_2}{c_1} \vee 1$, we have

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

which is precisely the desired result. $\square$

C Miscellaneous

C.1 Minimax

Lemma 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, $\Gamma_{\mathcal{H}} \leq \mu_{\nu_{\mathcal{H}}-1}$.

Proof. Consider that

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

Here, we have used the ordering of the eigenvalues and the fact that $\mu_{\nu_{\mathcal{H}}} \leq \frac{\sqrt{\nu_{\mathcal{H}} \log \left(1 + \frac{p}{s^2}\right)}}{n}$ to obtain the second term. The bound $\Gamma_{\mathcal{H}} \leq \mu_{\nu_{\mathcal{H}}-1}$ follows immediately from $\mu_{\nu_{\mathcal{H}}} \leq \mu_{\nu_{\mathcal{H}}-1}$ and $\frac{\sqrt{(\nu_{\mathcal{H}}-1) \log \left(1 + \frac{p}{s^2}\right)}}{n} \leq \mu_{\nu_{\mathcal{H}}-1}$ by definition of $\nu_{\mathcal{H}}$. $\square$

Proof of Lemma 1. The result is a special case of Lemma 2. $\square$

C.2 Adaptation

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

Proof. Consider that

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

Here, we have used the ordering of the eigenvalues and the fact that $\mu_{\nu_{\mathcal{H}}(s, a)} \leq \frac{\sqrt{\nu_{\mathcal{H}}(s, a) \log \left(1 + \frac{pa}{s^2}\right)}}{n}$ to obtain the second term. To bound $\Gamma_{\mathcal{H}} \leq \mu_{\nu_{\mathcal{H}}(s, a)-1}$ follows immediately from $\mu_{\nu_{\mathcal{H}}} \leq \mu_{\nu_{\mathcal{H}}(s, a)-1}$ and $\frac{\sqrt{(\nu_{\mathcal{H}}(s, a)-1) \log \left(1 + \frac{pa}{s^2}\right)}}{n} \leq \mu_{\nu_{\mathcal{H}}(s, a)-1}$ by definition of $\nu_{\mathcal{H}}(s, a)$. $\square$

Proof of Lemma 2. We first prove the second inequality. Since $\log \left(1 + \frac{pa}{s^2}\right) \leq \frac{n}{2}$ and $\mu_1 = 1$, it

immediately follows that $\nu_{\mathcal{H}}(s, a) \geq 2$. Therefore,

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

as desired.

Now we prove the first inequality. For any $\nu \geq \nu_{\mathcal{H}}(s, a)$, we have by the ordering of the eigenvalues

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

For any $\nu < \nu_{\mathcal{H}}(s, a)$, by definition of $\nu_{\mathcal{H}}(s, a)$ it follows that

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

Since for all ν we have shown

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

taking max over ν yields

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

as desired. $\square$

Proof of Lemma 3. For ease of notation, let us write $a = \mathcal{A}_{\mathcal{H}}$. By the assumption $\log(1 + pa) \leq \frac{n}{2}$ and $\mu_1 = 1$ we have $\nu_{\mathcal{H}}(s, a) \geq 2$ for all s . By definition of $\nu_{\mathcal{H}}(s, a)$, we have

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

and

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

Define

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

It is clear $\delta_s > 0$ for each s since the inequality in the penultimate display is strict. Moreover, observe that $\nu_{\mathcal{H}}(s, a + \delta) > \nu_{\mathcal{H}}(s, a) - 1$ for all $\delta < \delta_s$ by definition of $\nu_{\mathcal{H}}$. Moreover, since $\nu_{\mathcal{H}}$ is decreasing in its second argument, it thus follows that $\nu_{\mathcal{H}}(s, a + \delta) = \nu_{\mathcal{H}}(s, a)$ for all $\delta < \delta_s$. Letting $\delta^* := 1 \wedge \min_{s \in [p]} \delta_s$, it immediately follows that $\mathcal{V}_a = \mathcal{V}_{a + \delta}$ for all $\delta < \delta^*$. By definition of $\mathcal{A}_{\mathcal{H}}$, it follows that for any $\delta < \delta^*$

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

where we have used $\mathcal{A}_{\mathcal{H}} \geq 1$ and $\delta \leq 1$ to obtain the final inequality. The proof is complete. $\square$

C.3 Technical

Proposition 5 (Ingster-Suslina method [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

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

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

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

Lemma 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 $0 \leq k \leq s \wedge t$, then $E(\exp(\lambda Y)) \leq \left(1 - \frac{s}{n} + \frac{s}{n} e^\lambda\right)^t$ for $\lambda > 0$.

Corollary 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 $0 \leq k \leq s \leq n$, then $E(Y) = \frac{s^2}{n}$ and $E(\exp(\lambda Y)) \leq \left(1 - \frac{s}{n} + \frac{s}{n} e^\lambda\right)^s$ for $\lambda > 0$.