

SHARP MULTIPLE TESTING BOUNDARY FOR SPARSE SEQUENCES

BY KWEKU ABRAHAM^{1,a}, ISMAËL CASTILLO^{2,b} AND ÉTIENNE ROQUAIN^{2,c}

¹Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, alkwbraham@statslab.cam.ac.uk

²Laboratoire de Probabilités, Statistique et Modélisation (LPSM), Sorbonne Université, ismael.castillo@upmc.fr,
etienne.roquain@upmc.fr

This work investigates multiple testing by considering minimax separation rates in the sparse sequence model, when the testing risk is measured as the sum FDR+FNR (False Discovery Rate plus False Negative Rate). First, using the popular beta-min separation condition, with all nonzero signals separated from 0 by at least some amount, we determine the sharp minimax testing risk asymptotically and thereby explicitly describe the transition from “achievable multiple testing with vanishing risk” to “impossible multiple testing.” Adaptive multiple testing procedures achieving the corresponding optimal boundary are provided: the Benjamini–Hochberg procedure with a properly tuned level, and an empirical Bayes ℓ -value (“local FDR”) procedure. We prove that the FDR and FNR make nonsymmetric contributions to the testing risk for most optimal procedures, the FNR part being dominant at the boundary. The multiple testing hardness is then investigated for classes of *arbitrary* sparse signals. A number of extensions, including results for classification losses and convergence rates in the case of large signals, are also investigated.

1. Introduction.

1.1. *Background.* Multiple testing is a prominent topic of contemporary statistics, with a wide spectrum of applications including, for example, molecular biology, neuro-imaging and astrophysics. In this framework, many individual tests have to be performed simultaneously while controlling global error rates that take into account the multiplicity of the tests. A primary aim is to build procedures that guarantee control of a form of type I error, the most popular being the False Discovery Rate (FDR; see (5) below). For instance, the celebrated Benjamini–Hochberg (BH) procedure controls the FDR under independence [9]. Given an FDR controlling procedure, one may then ask whether it has a controlled type II error (or, equivalently, good power), measured for instance by the False Negative Rate (FNR; see (6) below).

In this context, a natural question is that of *optimality*: what is the best sum of type I and type II errors that is achievable by any multiple testing procedure? From a testing perspective and when a single test is considered, this can be answered via minimax separation rates between the two considered hypotheses, which have been investigated for a variety of nonparametric (see, e.g., [32], and [26] for an overview and further references) and high-dimensional models and loss functions; see, for example, [7, 31, 39]. The analogous question for multiple testing has received attention only very recently: the case of a familywise error risk is studied in [24], and in the case of FDR and FNR risks for deterministic sparse signals, aspects of this problem have been investigated in [5] and subsequently also in [8, 19, 40, 41]. More precise connections to these works are made below; see Section 1.7.

Received August 2023; revised March 2024.

MSC2020 subject classifications. 62G20, 62G15.

Key words and phrases. Multiple testing, sharp asymptotic minimaxity, False Discovery Rate, Benjamini–Hochberg procedure, frequentist analysis of Bayesian procedures.

1.2. *Sparse sequence model.* For some $\theta = (\theta_1, \dots, \theta_n)$ consider observing independent data $X = (X_1, \dots, X_n)$ satisfying, for a family of density functions $(f_a : a \in \mathbb{R})$,

$$(1) \quad X_i \sim f_{\theta_i}, \quad i = 1, \dots, n.$$

The vector θ is assumed to be sparse; that is, to belong to the set

$$(2) \quad \ell_0[s_n] = \{\theta \in \mathbb{R}^n, \|\theta\|_0 \leq s_n\}, \quad \|\theta\|_0 := \#\{1 \leq i \leq n : \theta_i \neq 0\},$$

consisting of vectors that have at most s_n nonzero coordinates, where $0 \leq s_n \leq n$; throughout the paper we consider the sparse asymptotic setting where

$$(3) \quad n \rightarrow \infty, \quad s_n \rightarrow \infty \quad \text{and} \quad n/s_n \rightarrow \infty.$$

We write P_θ for the law of X with parameter θ in (1) and E_θ for the corresponding expectation. Write $F_a(x) = \int_{-\infty}^x f_a(t) dt$ and $\bar{F}_a(x) = 1 - F_a(x)$ for the distribution function and the tail function, respectively. We will later place some regularity conditions on the F_a and some signal strength conditions on the θ .

EXAMPLE 1. The prototypical example to have in mind is the Gaussian location model, under which f_a is the density of the distribution $\mathcal{N}(a, 1)$, so that

$$(4) \quad X_i = \theta_i + \varepsilon_i, \quad \varepsilon_i \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1), \quad i = 1, \dots, n.$$

We first present our results in this model in Section 2, and then in Section 3 we give conditions for the more general model (1) under which corresponding results hold, allowing for a diverse range of models including for example Gaussian *scale* models.

The multiple testing problem consists of testing simultaneously, for each $1 \leq i \leq n$, the null hypothesis that there is no signal against the alternative hypothesis:

$$H_{0,i} : \text{“}\theta_i = 0\text{”} \quad \text{vs.} \quad H_{1,i} : \text{“}\theta_i \neq 0\text{”}.$$

1.3. *Multiple testing risks.* A multiple testing procedure is formally defined as a measurable function of the data $\varphi : x \in \mathbb{R}^n \mapsto (\varphi_i(x))_{1 \leq i \leq n} \in \{0, 1\}^n$, where, by convention, $\varphi_i(X) = 1$ corresponds to rejecting the null $H_{0,i}$. As such, the procedure will depend on n , and with some slight abuse of terminology, when dealing with asymptotics in terms of n , a sequence of such procedures is sometimes simply referred to as a “procedure” for short.

For any $\theta \in \mathbb{R}^n$ and any procedure φ , the false discovery rate (FDR) and the false discovery proportion (FDP) of φ at the parameter θ are respectively, defined as

$$(5) \quad \text{FDR}(\theta, \varphi) = E_\theta[\text{FDP}(\theta, \varphi)], \quad \text{FDP}(\theta, \varphi) = \frac{\sum_{i=1}^n \mathbf{1}\{\theta_i = 0\} \varphi_i(X)}{1 \vee \sum_{i=1}^n \varphi_i(X)}.$$

The false negative rate (FNR) at θ is here defined as (see, e.g., [5])

$$(6) \quad \text{FNR}(\theta, \varphi) = E_\theta \left[\frac{\sum_{i=1}^n \mathbf{1}\{\theta_i \neq 0\} (1 - \varphi_i(X))}{1 \vee \sum_{i=1}^n \mathbf{1}\{\theta_i \neq 0\}} \right].$$

The (multiple testing) combined risk at $\theta \in \mathbb{R}^n$ of a procedure φ is the sum

$$\mathfrak{R}(\theta, \varphi) = \text{FDR}(\theta, \varphi) + \text{FNR}(\theta, \varphi).$$

Given the popularity of FDR and FNR, this can be considered as a canonical notion of testing risk in the multiple testing context: it has indeed been considered for example, in [5] and the papers mentioned above. While the above is the main notion of risk used in this paper, other choices, including the classification risk, are discussed in Section 5.

1.4. *Separation of hypotheses and minimax testing risk.* To investigate questions of optimality, a natural benchmark is the minimax multiple testing risk, defined as

$$(7) \quad \mathfrak{R}(\Theta) = \inf_{\varphi} \sup_{\theta \in \Theta} \mathfrak{R}(\theta, \varphi),$$

where the infimum is over all multiple testing procedures and the parameter set Θ is some appropriate subset of $\ell_0[s_n]$. Typically, interesting parameter sets Θ are given to be “as large as possible,” while keeping $\mathfrak{R}(\Theta)$ in (7) at least strictly smaller than 1. Noting that the risk of the trivial procedure $\varphi_i = 0$ for all i is equal to 1, the latter corresponds to parameter configurations for which “nontrivial multiple testing” is achievable.

To investigate such “separation” rates in the present high-dimensional setting, perhaps the most popular approach is via a “beta-min” condition (see, e.g., [12], Section 7.4) meaning that all nonzero signals are above a certain threshold value. For instance, this condition is used for the (related but different) task of consistent model selection for estimators such as the LASSO; see Section 1.7 for more detail on this, and on other losses.

To fix ideas, let us consider the collection $\Theta(M)$ of vectors $\theta \in \ell_0[s_n]$ with nonzero coordinates taking only one possible value $M = M(n, s_n)$, in the Gaussian sequence model of Example 1. It follows from results in [5] (Theorem 2 therein) that if

$$(8) \quad M > a\sqrt{2 \log(n/s_n)}$$

for some constant $a > 1$, then the BH procedure with appropriately vanishing parameter has a vanishing \mathfrak{R} -risk. In addition, it is proved that no *thresholding-type procedure* can have a nontrivial \mathfrak{R} -risk uniformly over $\Theta(M)$ if $a < 1$ (Theorem 1 therein). These results suggest that, at least for thresholding procedures, the boundary of possible multiple testing is “close to” the threshold $\sqrt{2 \log(n/s_n)}$, which we refer to as the “oracle threshold” in the sequel.

1.5. *Questions of interest.* The previous discussion raises the following questions:

- how does the minimax risk $\mathfrak{R}(\Theta)$ behave for separated alternatives (e.g., for $\Theta = \Theta(M)$ with M as in (8)) when the infimum in (7) is taken over all possible procedures, not only thresholding ones?
- since multiple testing is “easy” when M is large and impossible when M is too small, what is the precise (asymptotic) boundary of signal strength M for which $\mathfrak{R}(\Theta)$ goes from 0 to 1? In other words, can one describe the transition from 0 to 1 in \mathfrak{R} when M decreases? This requires investigating the sharp minimax risk $\mathfrak{R}(\Theta)$.
- are there procedures that achieve the minimax risk, at least asymptotically, without knowledge of the sparsity parameter s_n ?
- suppose that rather than the s_n nonzero coordinates all equalling one signal value M , they can be divided into two different values M_1 and M_2 (e.g., as in Figure 1, middle column). Is this an easier or more difficult multiple testing problem than the former?

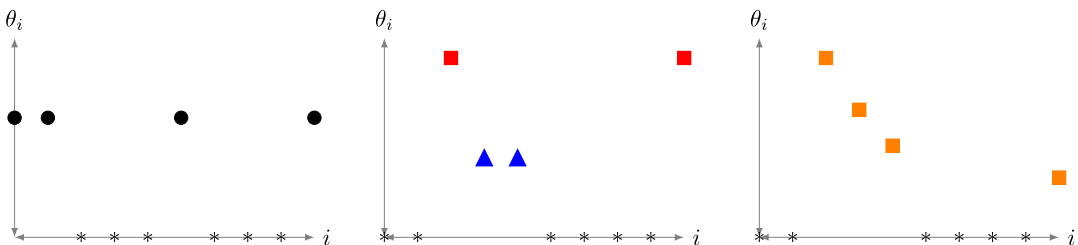


FIG. 1. Some signal configurations for nonzero θ_i 's to be considered. Left: equal signal strength; Middle: two distinct signals; Right: all signal values different. Zero θ_i 's are depicted by with a symbol $*$.

- (as asked by one referee:) under what assumptions on the noise can these questions can be addressed in a similar way?
- (as asked by one referee:) when M is large, for example, if $a > 1$ in (8), what is the optimal convergence rate of the risk to zero and can one find a procedure achieving this rate?

These and more general questions are addressed in the sequel.

As a specific aspect of testing problems in general, and of multiple testing in particular, it is common in practice to allow for a tolerance level, especially for the type I error, here typically for the FDR. Hence, for the combined risk under study here, it is not only the case where $\mathfrak{R}(\Theta) = o(1)$ that is of interest, but also the one where $\mathfrak{R}(\Theta)$ is of the order of a (possibly small) constant $c \in (0, 1)$. In addition, since the FDR and FNR account for errors that have different interpretations, it is also of interest to study the contribution of each error rate in the combined risk $\mathfrak{R}(\Theta)$. The results below will show that the contributions of FDR and FNR are not symmetric in this regime. Finally, we present our results asymptotically for simplicity—in particular this eases the presentation of results in the new setting of multiple signal strengths considered in Section 3—but the proofs can be adapted to give some nonasymptotic bounds.

1.6. Popular procedures: BH and empirical Bayes ℓ -values. Here, we describe two procedures that will be considered in the sequel.

First, probably the most widely used multiple testing procedure is the so-called Benjamini–Hochberg procedure, introduced in [9]. For some level α , it is given by $\varphi_\alpha^{\text{BH}} = (\mathbf{1}\{|X_i| \geq \hat{t}\})_{1 \leq i \leq n}$ where the threshold $\hat{t} = \hat{t}(\alpha)$ is defined as a specific intersection point between the empirical upper-tail distribution function of the X_i 's and a quantile curve of the noise distributions (see Section S-7 in [2] for details). To achieve good performances with respect to the combined risk, we will make use of the BH procedure where α is chosen to be slowly decreasing with n , as in [5, 11, 38]. A typical choice is $\alpha = \alpha_n \asymp 1/\sqrt{\log n}$.

The second procedure uses Bayesian ℓ -values (often also called local FDR values) with an empirical Bayes calibration. For a particular spike-and-slab prior Π_w on \mathbb{R}^n (see Section S-8 in [2] for details), we consider the empirical Bayes ℓ -value procedure defined by thresholding posterior probabilities of null hypotheses at some specified level $t \in (0, 1)$, that is, $\varphi_t^\ell = (\mathbf{1}\{\Pi_{\hat{w}}(\theta_i = 0|X) < t\})_{1 \leq i \leq n}$, where \hat{w} is the marginal maximum likelihood estimator for w , as in [34, 35]. The choice of t is not critical for obtaining a small combined risk, for example, $t = 0.3$ or $t = 1/2$ are possible choices. The widely used empirical Bayes spike-and-slab posterior distribution was investigated in terms of estimation properties in [18, 34], confidence sets in [20], and the resulting ℓ -value procedure was recently shown to control the FDR in [19]; see also [6] for an overview on the analysis of Bayesian high-dimensional posteriors and [1] for a related ℓ -value multiple testing algorithm.

1.7. Related literature, other modelling assumptions and risks. In the sparse sequence model, the interesting recent works [5, 41] provide bounds for the \mathfrak{R} -risk. In [5], the separation condition (8) is considered and the risk is shown to asymptotically converge to 0 for some procedures when $a > 1$, while it converges to 1 for any thresholding based procedure when $a < 1$. In [41], nonasymptotic lower bounds and upper bounds are further derived, for thresholding procedures, in the regime where $a = a_n > 1$ (possibly approaching 1) is known. These bounds are shown to be matched for the BH procedure with a suitably decreasing level. This analysis is further broadened and extended to more general models in the recent preprint [40] (see also Section S-13 in [2] for further discussion). In these works, which unlike the present work consider only thresholding procedures, the case where the risk converges to an arbitrary constant is not studied, hence the problem of identifying the sharp transition of the minimax risk from 0 to 1 was left open, as was the question of adapting to the signal strength for large signals.

In the multiple testing literature, a related way to measure optimality consists of finding a solution that minimizes the FNR while controlling a FDR-type error rate at level α ; see, for example, [22, 30, 42] in case of weighted procedures. This task is often done under the so-called “two-group mixture model,” introduced in [23], which assumes that each null hypothesis is true with some probability, and relies on specific ℓ -value (or local FDR) thresholding procedures; see [16, 17, 44, 45] and the recent work [29]. One way the present work differs from these references is in seeking not to minimise the FNR under a constraint, but rather to minimise the combined risk \mathfrak{R} . A more important difference is that we do not posit a mixture distribution for the true parameter θ , but rather assume it is deterministic and arbitrary (up to the sparsity constraint). Despite these differences, we will see that an ℓ -value thresholding based procedure (namely, $\varphi_i^{\hat{\ell}}$ as mentioned above) still achieves minimax performance.

Regarding related testing problems, the important work by Donoho and Jin [21] studies the detection problem for a single null and multiple alternatives; see also the subsequent works [4, 27, 28, 36]. In addition, the sparse sequence model has been much studied in terms of estimation for quadratic or ℓ^p -losses, here we only mention [3, 43] for their connections to multiple testing, where the authors use estimators related to the BH procedure.

Let us also mention that another FDR+FNR risk has been considered in [25], in a model with a single alternative distribution and using the knowledge of the number of true nulls. However, the FNR definition there is different from here: the denominator equals the number of accepted nulls, rather than the number of alternatives as here. In the case of sparse signal, these two FNR notions scale very differently; the former scaling is not well-suited to deal with sparsity, while using the FNR notion considered herein enables us to exhibit a sharp phase transition phenomenon. The recent work [8] derives some robust results for model selection based procedures, including for various notions of sums FDR+FNR, but these results are restricted to a range where the multiple testing risk tends to zero at a certain rate.

Finally, a different but somewhat related loss function is the Hamming or classification loss. The corresponding classification risk is considered in [11, 38] in a two-group mixture model, with lower bounds restricted to thresholding classifiers. There, it is proved that the BH procedure with a suitably vanishing level achieves the oracle performance. However, these results study the Bayes risk (i.e., the minimum average risk) in this mixture model and do not provide a complete minimax analysis. Further extensions are derived in [33], for example, by handling more general dependent models. Coming back to the Gaussian sequence model (4) (with nonrandom θ_i 's), minimax Hamming estimators are derived in [13, 14] (see also the earlier work [15]), where the authors study the boundary for exact recovery (the classification risk goes to 0) and almost sure recovery (the classification risk is a vanishing fraction of the sparsity parameter). We refer to Section 5 for more on the classification risk.

1.8. Outline and notation. Section 2 contains a first series of results in the setting of Example 1 under a “beta-min” condition on signals. This enables us to present some main results of the paper in a simple case first. Namely, the asymptotic minimax risk is computed and adaptive procedures achieving this risk are given, with the FNR shown to be the dominating term in the risk for most procedures. Section 3 states the main results under general conditions: the signal strengths are arbitrary and the model signal and noise distributions more general. Illustrations are provided in Section 4. Classification loss and some other risks are investigated in Section 5. Section 6 considers the case of large signals with risks rapidly decreasing toward zero and Section 7 concludes with a brief discussion. While one key pair of results is proved in Section 8, other proofs are postponed to the Supplementary Material [2] (references to which are given the prefix S-).

Notation. The density of a standard normal variable is denoted by ϕ and we write Φ for the cumulative distribution function $\Phi(x) = \int_{-\infty}^x \phi(t) dt$ and $\bar{\Phi} = 1 - \Phi$ for the tail probabilities.

We use $\underline{\lim}_{n \rightarrow \infty}$, $\overline{\lim}_{n \rightarrow \infty}$ to denote the liminf or the limsup, respectively. We use $x_n \lesssim y_n$, $y_n \gtrsim x_n$ or $x_n = O(y_n)$ to signify that there exists $C > 0$ such that $x_n \leq Cy_n$ for all n large; we write $x_n \asymp y_n$ if $x_n \lesssim y_n$ and $y_n \lesssim x_n$; we write $x_n \sim y_n$ if $x_n/y_n \rightarrow 1$; and we write $x_n \ll y_n$, $y_n \gg x_n$ or $x_n = o(y_n)$ if $x_n/y_n \rightarrow 0$. We use this notation correspondingly for functions $f(x)$, $g(x)$, with the limits taken either as $x \rightarrow \infty$ or as $x \rightarrow 0$ depending on context. A sequence of random variables is said to be $o_P(1)$ if it converges to 0 in probability under the data generating law P_θ . For reals a, b , we set $a \vee b = \max(a, b)$ and $a \wedge b = \min(a, b)$. Finally, for a finite set A , we denote by $|A|$ or $\#A$ its cardinality.

2. Sharp boundaries: Beta-min condition and Gaussian noise. In this section, we focus on the Gaussian location model presented in Example 1. To evaluate the minimax risk, we define a class of configurations for θ that measures how the alternatives are separated from the null hypothesis: for a given $a \in \mathbb{R}$, set

$$(9) \quad \Theta = \Theta(a, s_n) = \{\theta \in \ell_0[s_n] : |\theta_i| \geq a \text{ for } i \in S_\theta, |S_\theta| = s_n\},$$

where $S_\theta = \{i : \theta_i \neq 0\}$ denotes the support of θ (recall $\ell_0[s_n]$ is defined by (2)). This choice corresponds to a so-called beta-min type condition, meaning that all intensities of nonzero coefficients are required to be above a certain value $a > 0$. For example, each of the signals depicted in Figure 1 belongs to some class $\Theta(a, s_n)$; in the first panel, one may take a to be the shared value of the nonzero θ_i 's, whereas in the third panel a can at most be taken to be the smallest nonzero value. More general classes, leading to more refined results for the second and third panels in Figure 1, are considered in Section 3.

2.1. *Minimax multiple testing risk.* Previous works in the literature [5, 41] suggest that the phase transition for the combined risk over $\Theta(a, s_n)$ arises when a is close to the oracle threshold $\sqrt{2 \log(n/s_n)}$. We identify here a sharp formulation for this boundary, by considering for $b \in \mathbb{R}$,

$$(10) \quad \Theta_b = \Theta(a_b, s_n), \quad a_b = \sqrt{2 \log(n/s_n)} + b,$$

with $\Theta(a, s_n)$ as in (9). The following result holds.

THEOREM 1. *In the Gaussian location model of Example 1 consider $\Theta_b = \Theta(a_b, s_n)$ as in (10). For any fixed $b \in \mathbb{R}$, under the sparse asymptotics (3) the minimax \mathfrak{R} -risk over Θ_b verifies*

$$\inf_{\varphi} \sup_{\theta \in \Theta_b} \mathfrak{R}(\theta, \varphi) = \overline{\Phi}(b) + o(1).$$

The result also holds in the limiting cases $b = b_n \rightarrow +\infty$ and $b = b_n \rightarrow -\infty$.

Since $\overline{\Phi}(b)$ increases from 0 to 1 when b decreases from $+\infty$ to $-\infty$, Theorem 1 exhibits an asymptotic phase transition and shows that the considered boundary (10) is sharp. It follows from the proof of the result that the oracle thresholding rule

$$(11) \quad \varphi_i^* = \mathbf{1}\{|X_i| \geq \sqrt{2 \log(n/s_n)}\}, \quad 1 \leq i \leq n,$$

is asymptotically minimax, independent of the value of b .

2.2. *Applications: Nontrivial testing and conservative testing.* Let us now provide two consequences of Theorem 1 for $\Theta_b = \Theta(a_b, s_n)$ as in (10), for respectively “nontrivial” and “conservative” testing, as defined in the following statement.

COROLLARY 1. *Consider the Gaussian location model of Example 1.*

For any fixed $b \in \mathbb{R}$, asymptotic nontrivial testing is possible over Θ_b , in that there exists a procedure φ such that

$$\lim_{n \rightarrow \infty} \sup_{\theta \in \Theta_b} \mathfrak{R}(\theta, \varphi) < 1.$$

In contrast, for any sequence $b = b_n \rightarrow -\infty$, we have

$$\lim_{n \rightarrow \infty} \inf_{\varphi} \sup_{\theta \in \Theta_b} \mathfrak{R}(\theta, \varphi) = 1,$$

that is, asymptotic nontrivial testing is impossible.

If $b = b_n \rightarrow +\infty$, asymptotic conservative testing is possible over Θ_b , in that there exists a procedure φ such that

$$\lim_{n \rightarrow \infty} \sup_{\theta \in \Theta_b} \mathfrak{R}(\theta, \varphi) = 0.$$

In contrast, for any fixed $b \in \mathbb{R}$, we have

$$\lim_{n \rightarrow \infty} \inf_{\varphi} \sup_{\theta \in \Theta_b} \mathfrak{R}(\theta, \varphi) > 0,$$

that is, asymptotic conservative testing is impossible over Θ_b for any procedure.

We identify a new sharp boundary for asymptotic *nontrivial* testing: $a_b = \sqrt{2 \log(n/s_n)} + b$ with $b \rightarrow -\infty$ corresponds to the regime where it is impossible to build a multiple testing procedure doing (asymptotically) better than the trivial ones $\varphi_i = 0$ for all i or $\varphi_i = 1$ for all i . On the contrary, provided that b is a finite constant (which may be negative), some nontrivial control of the \mathfrak{R} -risk is possible.

Further, we find the following sharp boundary for asymptotic *conservative* testing: it is possible when $a_b = \sqrt{2 \log(n/s_n)} + b$ with $b = b_n \rightarrow +\infty$, but becomes impossible when b is finite. It is interesting to note that a similar boundary with $b = b_n \rightarrow +\infty$ was identified in [14] for the so-called *almost-full recovery* problem when the risk is the Hamming loss and the goal is to correctly classify up to a $o(s_n)$ number of nonzero signals. In this view, the second part of Corollary 1 can be seen as an analogue of Theorem 4.3 in [14] for the multiple testing risk. We refer to Section 5 (and Section S-3 in [2]) for a more detailed connection to the classification problem.

2.3. *Adaptation.* The asymptotically minimax procedure (11) requires the knowledge of the sparsity parameter s_n . A natural question is whether there exists an *adaptive* multiple testing procedure that achieves the minimax risk without using the knowledge of s_n . Such procedures do exist. Some will require an additional *polynomial sparsity* assumption: for some unknown $c < 1$,

$$(12) \quad s_n \lesssim n^c.$$

The empirical Bayes and BH procedures in the following statement are defined precisely in the Supplementary Material [2]; see, respectively, (S-29) and (S-18) therein.

THEOREM 2. *In the Gaussian location model of Example 1, consider $\Theta_b = \Theta(a_b, s_n)$ as in (10) where b is either fixed in \mathbb{R} or is a sequence $b = b_n$ tending to $+\infty$. There exists a multiple testing procedure φ , not depending on s_n or b , such that in the sparse asymptotics (3)*

$$\sup_{\theta \in \Theta_b} \mathfrak{R}(\theta, \varphi) = \overline{\Phi}(b) + o(1).$$

In particular this holds for the empirical Bayes ℓ -value procedure (S-29) in [2] taken at any fixed threshold $t \in (0, 1)$. In addition, under (12), this also holds for the BH procedure (S-18) in [2] taken at a vanishing level $\alpha = \alpha_n = o(1)$ that satisfies $-\log(\alpha_n) = o(\sqrt{\log n})$.

Theorem 2 establishes that two popular (and simple) procedures are asymptotically sharp minimax adaptive. In particular, they automatically achieve the nontrivial and conservative boundaries described in Corollary 1. One advantage of the empirical Bayes ℓ -value procedure over the BH procedure is that it does not need any further parameter tuning in order to be valid. Nevertheless, an advantage of the BH procedure is that its FDR is always equal to $(1 - |S_\theta|/n)\alpha$ (see [10]) so that one has more explicit information about the first term of the combined risk.

REMARK 1 (Case $b = b_n \rightarrow -\infty$). When $b_n \rightarrow -\infty$, the lower bound of Theorem 1 shows that nontrivial testing is impossible. One could in principle have $\mathfrak{R}(\theta, \varphi) = 2 > 1$ so that the upper bounds in Theorem 2 are not automatic. However, the BH procedure at level $\alpha_n = o(1)$ achieves the limit $1 = \overline{\Phi}(-\infty)$ in this case, as a consequence of the explicit control of its FDR noted above. Under polynomial sparsity (12) the same is true of the ℓ -value procedure (with $t \leq 3/4$) as an immediate consequence of Theorem 1 in [19].

REMARK 2 (Polynomial sparsity assumption for the BH procedure). The condition $-\log(\alpha_n) = o(\sqrt{\log n})$ for the BH procedure is typically achieved by choosing $\alpha_n = 1/\sqrt{\log n}$. A rate of convergence of the risk to $\overline{\Phi}(b)$ of $(\log \log n)/\sqrt{\log n} + e^{-cs_n}$ can be obtained for such choice; see Theorem 5 and the paragraph thereafter. In addition, the polynomial sparsity assumption can be relaxed slightly; see Remark S-2 in [2] (second bullet).

2.4. Sparsity preserving procedures and dominating FNR at the boundary. Theorem 1 determines the minimax optimal combined risk \mathfrak{R} , that is, the sum FDR + FNR, but not the balance between the two terms struck by optimal procedures. In this section, we investigate this tradeoff in more details. For instance, the proof of Theorem 1 shows that the oracle procedure (11) achieves an optimal multiple testing risk by satisfying

$$(13) \quad \sup_{\theta \in \Theta_b} \text{FDR}(\theta, \varphi^*) = o(1), \quad \sup_{\theta \in \Theta_b} \text{FNR}(\theta, \varphi^*) = \overline{\Phi}(b) + o(1),$$

that is, the FNR “spends” all the allowed “budget” from the overall minimax risk. In this section, we show that this phenomenon holds true for most “reasonable” procedures. Clearly, some restriction is required on the class of procedures as the trivial test $\varphi \equiv 1$ that always rejects the null achieves the optimal asymptotic risk (of 1) over Θ_b for $b = b_n \rightarrow -\infty$ and has FNR zero. We see below that for any “sparsity preserving” procedure, that is, one not overshooting the true sparsity index s_n by more than some large multiplicative factor, the FNR alone cannot go below $\overline{\Phi}(b)$ asymptotically.

DEFINITION 1. We say that a multiple-testing procedure $\varphi = \varphi(X) \in \{0, 1\}^n$ (or strictly, a sequence of such procedures, indexed by n), is *sparsity-preserving over* $\Theta = (\Theta_n)_n$ (with

$\Theta_n \subset \ell_0[s_n]$) up to a multiplicative factor $A = (A_n)_n$ if, as $n \rightarrow \infty$,

$$(14) \quad \sup_{\theta \in \Theta_n} P_\theta \left[\sum_{i=1}^n \varphi_i(X) > A_n s_n \right] = o(1).$$

We denote by $\mathcal{S}_A(\Theta) = \mathcal{S}_A((s_n)_n, \Theta)$ the set of all such procedure sequences.

The sparsity preserving property entails a total number of rejections not exceeding $A_n s_n$ with high probability, and can be interpreted as a weak notion of type I error rate control. Many procedures (or also estimators) encountered in the literature on sparse classes verify this condition for A_n either a large enough constant or going to infinity slowly, even when taking the supremum in (14) over the whole sparse class $\ell_0[s_n]$. Several examples are provided in Section S-2 in [2], including the BH procedure (with fixed level $\alpha < 1$ or with $\alpha = \alpha_n \rightarrow 0$), the ℓ -value procedure, the oracle thresholding procedure used to prove Theorem 1, and, more generally, procedures controlling the FDP (5) in a specific sense.

THEOREM 3. *In the setting of Theorem 1, consider a fixed real b , a sequence $B = (B_n)_n$ with $B_n^2 \leq e^{(\log(n/s_n))^{1/4}}$ and $\underline{\lim}_n B_n > 1$, and consider $\mathcal{S}_B(\Theta_b)$ as in Definition 1. Then we have in the sparse asymptotics (3)*

$$(15) \quad \inf_{\varphi \in \mathcal{S}_B(\Theta_b)} \sup_{\theta \in \Theta_b} \text{FNR}(\theta, \varphi) = \overline{\Phi}(b) + o(1).$$

The result continues to hold if $b = b_n \rightarrow +\infty$ or $b = b_n \rightarrow -\infty$.

Since Theorem 2 shows that the (sparsity preserving) ℓ -value and BH procedures have their FNR suitably upper bounded, these procedures both achieve the bound (15), under polynomial sparsity (12) for the BH procedure (note that the bound (15) trivially holds for any procedure in the case $b = b_n \rightarrow -\infty$).

Theorem 3 sharpens the lower bounds of Theorem 1 by stating that the combined risk \mathfrak{R} can in fact be replaced by the FNR (i.e., the type II error) only, if one is willing to restrict slightly the class of multiple testing procedures to φ 's that do not often reject more than $B_n s_n$ hypotheses, where B_n is permitted to diverge at some rate specified above. Some intuition behind this result is given in Section 4 (in particular, see the discussion of Figure 2).

This phenomenon of dominating FNR is a novel finding. We would like to underline two related points. First, it is specific to the regime where the \mathfrak{R} -risk is bounded away from 0 (otherwise the lower bound in (15) is trivial); in regimes with fairly strong signals, such as ones studied in Section 6, both FDR and FNR typically vary on the same level. Second, it qualitatively explains results in Section 5, where we will show that the sharp minimax constant for the normalised classification risk is the same as for the \mathfrak{R} -risk. Indeed, both risks have the same type II-error risk (equal to the FNR).

A straightforward but interesting consequence of Theorem 3 is that no sparsity preserving procedure can “trade” some loss in FDR for some improvement of the FNR while still staying close to the optimal risk: if its FDR is equal to $\alpha \in (0, 1)$ asymptotically (such as the standard BH procedure for a fixed level α), it must miss the sharp combined risk \mathfrak{R} over Θ_b by at least an additive factor α . Another (surprising) consequence of Theorem 3 for “top- K ” procedures is given in Section S-9.2 in [2]; see Corollary S-3 therein.

COROLLARY 2. *In the setting of Theorem 3, for some $\alpha \in (0, 1)$ and $b \in \mathbb{R}$, let φ_α be any sparsity preserving procedure $\varphi_\alpha \in \mathcal{S}_B(\Theta_b)$ such that $\underline{\lim}_n \inf_{\theta \in \Theta_b} \text{FDR}(\theta, \varphi_\alpha) \geq \alpha > 0$. Then φ_α must miss the asymptotically minimax combined risk by at least α , that is,*

$$\underline{\lim}_n \sup_{\theta \in \Theta_b} \mathfrak{R}(\theta, \varphi_\alpha) \geq \alpha + \overline{\Phi}(b).$$

This is in particular the case of the BH procedure with fixed level α , as defined in (S-18) in [2].

Corollary 2 follows from Theorem 3, the fact that the BH procedure with fixed parameter is sparsity preserving (Section S-2 in [2]) and the explicit expression for the FDR of the BH procedure; see (S-19) in [2].

Finally, for completeness, let us mention that for a procedure which is not sparsity preserving, trading FDR for FNR is formally possible. A (somewhat degenerate) example is given in Section S-2 in [2]; see Example S-1 therein.

3. Sharp boundaries: Arbitrary signal strengths and beyond Gaussian noise. This section presents more refined results, by relaxing the assumptions on both the signal strength and the noise. These results show exactly how the limiting risk depends on a measure of the signal strengths, in a general sequence model which allows for nonadditive and non-Gaussian noise. This generalisation also sheds light on the previous Gaussian results, by revealing an underlying sufficient set of assumptions for the proofs. In addition, we are also able to prove local, “pointwise,” versions of the results, that we largely postpone to Section S-4 in [2] for clarity of presentation.

3.1. *Extended noise assumption.* Recall from (1)–(3) that we assume $X_i \sim f_{\theta_i}$, $i \leq n$, independently, with $(f_a : a \in \mathbb{R})$ a family of densities and $\theta \in \ell_0[s_n]$, with $s_n \rightarrow \infty$ and $n/s_n \rightarrow \infty$. Recall the notation F_a and \overline{F}_a therein.

ASSUMPTION 1. There exists a constant L such that each F_a is L -Lipschitz. There exist sequences of positive numbers $a_n^* \rightarrow \infty$, $\delta_n \rightarrow 0$ such that

$$(16) \quad (n/s_n)\overline{F}_0(a_n^* - \delta_n) \rightarrow \infty,$$

$$(17) \quad (n/s_n)\overline{F}_0(a_n^*) \rightarrow 0.$$

The density f_0 is continuous and positive on \mathbb{R} . Further assume one of:

A. $f_{-a}(-x) = f_a(x)$ for $a, x \in \mathbb{R}$, $\overline{F}_a(x)$ is increasing in $a \in \mathbb{R}$ and, for $a > 0$,

$$(18) \quad f_a(x)/f_0(x) \text{ is increasing in } x \in \mathbb{R}.$$

B. $f_a(-x) = f_a(x)$ for all $a, x \in \mathbb{R}$, $\overline{F}_a(x)$ is increasing in $a > 0$ for $x > 0$, and for $a \neq 0$

$$(19) \quad f_a(x)/f_0(x) \text{ is increasing in } x > 0.$$

Assumption 1 was partly inspired by [40]. It is explored in depth in Section S-1 in [2] and we report essential facts below.

REMARK 3. Assumption 1A is designed for location models: the symmetry property of the densities, and monotonicity and Lipschitz continuity of the distribution functions are automatic in a symmetric location model where $f_a(x) = f_0(x - a)$, $f_0(x) = f_0(-x)$ for some bounded positive density f_0 . Assumption 1B is designed for scale models: the symmetry property of the densities and monotonicity and Lipschitz continuity of the distribution functions are again automatic if $f_a(x) = f_0(x/\sqrt{1 + |a|})$ and $f_0(x) = f_0(-x)$ for some bounded positive density f_0 . Next, (16)–(17) are conditions on the “null” distribution \overline{F}_0 that allow for a sharp description of the minimax risk on the boundary, as in the previous Gaussian case. More precisely, (16) means that the expected number of nulls larger than $a_n^* - \delta_n$ is much larger than s_n , while (17) means that the expected number of nulls larger than a_n^* is $o(s_n)$. Hence, up to a concentration argument, and assuming that a nontrivial portion of the s_n true

signals can be recovered, (16) and (17) entail that the FDR of a thresholding based procedure $\varphi_t = \mathbf{1}_{|X_i| \geq t}$ has a sharp decay from 1 to 0 when t increases from $a_n^* - \delta_n$ to a_n^* . Thus, under Assumption 1, the optimal threshold should be at $a_n^* + o(1)$. This sharp transition is an essential phenomenon in this setting; it will be further illustrated in Section 4.1.

We now give two illustrative examples of settings in which Assumption 1 holds.

EXAMPLE 2. Assumption 1A holds with $a_n^* = (\zeta \log(n/s_n))^{1/\zeta}$ for the Subbotin (generalised Gaussian) location model, with f_a denoting the law of $X = a + \varepsilon$, $\varepsilon \sim \phi_\zeta$,

$$(20) \quad \phi_\zeta(x) = L_\zeta^{-1} e^{-|x|^\zeta/\zeta}, \quad \zeta > 1;$$

see Lemma S-1 in [2]. In particular, this model includes standard Gaussian noise as the case $\zeta = 2$, while the excluded case $\zeta = 1$ would correspond to Laplace noise. In this setting, we will write $\overline{\Phi}_\zeta(x) = \int_x^\infty \phi_\zeta(t) dt$ and $\Phi_\zeta = 1 - \overline{\Phi}_\zeta$. Note that when we write a_n^* in reference to the Subbotin model, we will mean the value $(\zeta \log(n/s_n))^{1/\zeta}$ except where specified otherwise.

EXAMPLE 3. Assumption 1B holds with $a_n^* = (\zeta \log(n/s_n))^{1/\zeta}$ for the Subbotin scale model, with f_a denoting the law of $X = (1 + |a|)^{1/2} \varepsilon$, $\varepsilon \sim \phi_\zeta$, with ϕ_ζ given by (20); see Lemma S-1 in [2].

In this section, we also consider a more general parameter set, with different signal strengths. For a vector $\mathbf{a} = (a_1, \dots, a_{s_n})$ (implicitly indexed by n) with $a_j > 0$ for $1 \leq j \leq s_n$, define

$$(21) \quad \Theta(\mathbf{a}, s_n) = \{\theta \in \ell_0[s_n] : \exists i_1, \dots, i_{s_n} \text{ all distinct, } |\theta_{i_j}| \geq a_j, 1 \leq j \leq s_n\}.$$

Note this generalises the definition (9): we recover $\Theta(\mathbf{a}, s_n) = \Theta(a, s_n)$ if $\mathbf{a} = (a, a, \dots, a)$. In general, we have only the inclusion $\Theta(\mathbf{a}, s_n) \subseteq \Theta(\min_{1 \leq j \leq s_n} a_j, s_n)$, so that this definition allows for considering a more subtle minimax tradeoff. A key quantity is

$$(22) \quad \Lambda_n(\mathbf{a}) = s_n^{-1} \sum_{j=1}^{s_n} F_{a_j}(a_n^*);$$

this can be seen roughly as a measure of the (lack of) signal strength for the parameter set $\Theta(\mathbf{a}, s_n)$. In a slight abuse of notation, if θ is a vector whose nonzero entries have absolute values a_1, \dots, a_{s_n} and recalling the notation $S_\theta = \{i : \theta_i \neq 0\}$, we also write

$$(23) \quad \Lambda_n(\theta) = s_n^{-1} \sum_{i \in S_\theta} F_{|\theta_i|}(a_n^*).$$

We show in Theorem 4 below that the asymptotic behaviour of $\Lambda_n(\mathbf{a})$ drives the minimax risk over the class $\Theta(\mathbf{a}, s_n)$. We now give three examples of location models to illustrate how different signal shapes and values affect $\Lambda_n(\mathbf{a})$.

EXAMPLE 4 (Location model with single signal strength close to a_n^*). A simple case is the Subbotin location model of Example 2, with $\theta_i = a_n^* + b$ for all $i \in S_\theta$, for some fixed $b \in \mathbb{R}$. Then θ lies on the boundary of the beta-min parameter set and Λ_n in (23) reads $\Lambda_n(\theta) = \overline{\Phi}_\zeta(b)$. This corresponds to the first case in Figure 1.

EXAMPLE 5 (Location model with two signal strengths near a_n^*). For fixed $x, y \in \mathbb{R}^2$, let $M = \max(x, y)$, $m = \min(x, y)$. Consider the Subbotin location model of Example 2, with $\theta_i = a_n^* + M$ for $\lfloor s_n \beta \rfloor$ coefficients in S_θ and $\theta_i = a_n^* + m$ for the other coefficients in S_θ , where $\beta \in (0, 1)$ is a given proportion of stronger signal. Then inserting $F_a = \Phi_\zeta(\cdot - a)$ in (23) we obtain

$$\begin{aligned} \Lambda_n(\theta) &= \frac{\lfloor s_n \beta \rfloor}{s_n} \overline{\Phi}_\zeta(M) + \frac{s_n - \lfloor s_n \beta \rfloor}{s_n} \overline{\Phi}_\zeta(m) \\ &\rightarrow \beta \overline{\Phi}_\zeta(M) + (1 - \beta) \overline{\Phi}_\zeta(m) =: \Lambda_\infty, \end{aligned}$$

as $n \rightarrow \infty$. This corresponds to the second picture of Figure 1. Level sets of $\Lambda_\infty = \Lambda_\infty(x, y)$ for this θ are displayed in Figure 3 below.

In the previous examples, we have taken signals around the critical threshold a_n^* . We can allow more generally for arbitrary nonzero signals.

EXAMPLE 6 (Gaussian location model, mixed signal). Consider to fix ideas the Gaussian model of Example 1 and suppose the nonzero entries of θ are given by $(Ai/s_n)\sqrt{2\log(n/s_n)}$ for $i = 1, \dots, s_n$ and $A \geq 1$ fixed. This is a special example of the third case in Figure 1. Then

$$\Lambda_n(\theta) = A^{-1} + o(1)$$

(to check this, one can, e.g., separate signal coordinates i 's into three subsets delimited by $(s_n/A)(1 \pm r_n)$ for r_n tending to zero suitably slowly). In this example, what contributes to $\Lambda_n(\theta)$ is the proportion of signals below the (asymptotically) optimal threshold $a_n^* = \sqrt{2\log(n/s_n)}$.

3.2. *Main results in the general setting.* We extend Theorems 1, 2 and 3, respectively, to the more general noise and parameter set introduced in this section.

THEOREM 4. Consider the sparse sequence model (1)–(3) with Assumption 1. Consider a vector $\mathbf{a} = (a_1, \dots, a_{s_n}) \in \mathbb{R}_+^{s_n}$, $\Theta(\mathbf{a}, s_n)$ defined by (21) and $\Lambda_n(\mathbf{a})$ defined by (22). Then:

- Under Assumption 1A,

$$\inf_{\varphi} \sup_{\theta \in \Theta(\mathbf{a}, s_n)} \mathfrak{R}(\theta, \varphi) = \Lambda_n(\mathbf{a}) + o(1);$$

- Under Assumption 1B,

$$\inf_{\varphi} \sup_{\theta \in \Theta(\mathbf{a}, s_n)} \mathfrak{R}(\theta, \varphi) = 2\Lambda_n(\mathbf{a}) - 1 + o(1).$$

In each case the risk bound is achieved by a thresholding procedure $\varphi_i(X) = \mathbf{1}\{|X_i| > a_n^*\}$ with a_n^* as in Assumption 1.

Theorem 4 in particular applies in the Subbotin location model of Example 2 with $\overline{F}_0 = \overline{\Phi}_\zeta$, $a_n^* = (\zeta \log(n/s_n))^{1/\zeta}$ and a suitable choice of δ_n ; see Lemma S-1 in [2].

To prove this theorem, an easy heuristic can be built from Remark 3, since by the reasoning therein only a threshold $t = a_n^* + o(1)$ leads to nontrivial FDR and the FNR for that threshold is precisely of order $\Lambda_n(\mathbf{a})$ and $2\Lambda_n(\mathbf{a}) - 1$ (under Assumption 1A and 1B, respectively). However, the rigorous proof is significantly more involved; in particular, the lower bound argument is more sophisticated, because we do not restrict to thresholding procedures; see Section 8.

Recall our motivating question from Figure 1, of whether all nonzero signals being equal (left panel) is easier or harder for multiple testing than half taking one value and half another value (middle panel). Applying Theorem 4 in Example 5, the asymptotic difficulty of testing is governed by the quantity $\Lambda_\infty = \Lambda_\infty(x, y)$. See Section 4.1 for some examples of how Λ_∞ varies with x and y , but note immediately that Λ_∞ is smaller than the bound $\overline{\Phi}_\zeta(m)$ attainable using that $\Theta((a_n^* + m, \dots, a_n^* + M, \dots), s_n)$ is a subset of $\Theta(a_n^* + m, s_n)$.

The procedure exhibited to prove the upper bound in Theorem 4, like that of Theorem 1, is an ‘‘oracle’’ procedure requiring knowledge of the a_n^* , which typically depends on the sparsity s_n . The next result gives an upper bound on the pointwise and uniform risk of the Benjamini–Hochberg procedure under Assumption 1; this will yield adaptivity to s_n in many settings. Let us introduce the following condition on the level $\alpha = \alpha_n$ of the BH procedure:

$$(24) \quad \frac{3n\overline{F}_0(a_n^*)}{s_n(1 - \Lambda_n(\theta))} \leq \alpha_n \leq \min\left\{1, \frac{n\overline{F}_0(a_n^* - \delta_n)}{s_n}\right\} \quad \text{for } n \text{ large enough.}$$

THEOREM 5. *Consider the sparse sequence model (1)–(3) and grant Assumption 1A. Then the BH procedure taken at a level $\alpha = \alpha_n$ obeying (24) satisfies, for all $\theta \in \mathbb{R}^n$ with $|\mathcal{S}_\theta| = s_n$, for n large enough,*

$$(25) \quad \mathfrak{R}(\theta, \varphi) \leq \Lambda_n(\theta) + \alpha_n + \exp\left(-\frac{(1 - \Lambda_n(\theta))^2}{32}s_n\right).$$

It follows that the BH procedure achieves the bound of Theorem 4,

$$\sup_{\theta \in \Theta(\mathbf{a}, s_n)} \mathfrak{R}(\theta, \varphi) \leq \Lambda_n(\mathbf{a}) + o(1).$$

In addition, if $\Lambda_n(\mathbf{a})$ is bounded away from 1 we have the concrete expression $\alpha_n + \exp(-(1 - \Lambda_n(\mathbf{a}))^2 s_n / 32)$ for the $o(1)$ terms. Finally, if we instead grant Assumption 1B, the same bounds hold with $\Lambda_n(\cdot)$ replaced by $2\Lambda_n(\cdot) - 1$.

Theorem 5 is proved in Section S-7.2 in [2]. It implies that the risk bound of Theorem 4 can be attained adaptively to s_n in any case where a valid level α_n can be chosen. Typically, the condition (24) is achieved for $\Lambda_n(\theta)$ bounded away from 1 by choosing $\alpha_n \rightarrow 0$ with $(n/s_n)\overline{F}_0(a_n^*)/\alpha_n \rightarrow 0$ (and noting that $(n/s_n)\overline{F}_0(a_n^* - \delta_n) \geq 1$ for n large). For instance, in the Subbotin case with parameter ζ , under the assumption of polynomial sparsity (12), condition (24) can be achieved if $\alpha_n = o(1)$ satisfies $\log(1/\alpha_n) = o((\log n)^{1-1/\zeta})$; see Lemma S-1 and Remark S-2 in [2]. In addition, choosing $\alpha_n \asymp (\log n)^{-(1-1/\zeta)}$ leads to the convergence rate $(\mathfrak{R}(\theta, \varphi) - s_n^{-1} \sum_{i \in \mathcal{S}_\theta} F_{|\theta_i|}((\zeta \log(n/s_n))^{1/\zeta})) \lesssim \frac{\log \log n}{(\log n)^{1-1/\zeta}} + e^{-cs_n}$, for some constant $c > 0$; see again Remark S-2. Note that this shares similarities with the rate obtained in [38] for the classification risk of BH procedure.

REMARK 4. The inequality in (25) is in fact an equality, up to $o(1)$ terms, so that the result quantifies fairly precisely the risk actually obtained by the Benjamini–Hochberg procedure in a sparse sequence model setting; see Remark S-4 after Theorem S-1 in [2].

Let us remark that the ℓ -value procedure also achieves the bound in the Gaussian case with multiple signal levels; see Section S-8.2 in [2]. We conjecture that the ℓ -value procedure also achieves the bound in the Subbotin model, and possibly even in the general noise model under Assumption 1, but proving this would require substantial extra technical work.

The following result, proved in Section 6, shows that the FNR is the dominating term at the boundary when focusing on sparsity preserving procedures even in the general framework of this section. Recall that the BH procedure is sparsity preserving (proved in Section S-2 in [2]), hence achieves the bounds of Theorem 6 adaptively under the conditions of Theorem 5.

THEOREM 6. *Consider the settings of Theorem 4. For any sequence $B = (B_n)_n$ with $B_n^2 \leq \frac{1}{3}(n/s_n)\overline{F}_0(a_n^* - \delta_n)$ and $\underline{\lim}_n B_n > 1$, let \mathcal{S}_B denote the set of sparsity preserving procedures over the set $\Theta = \Theta(\mathbf{a}, s_n)$, as in Definition 1. Then the conclusions of Theorem 4 hold with $\mathfrak{R}(\theta, \varphi)$ replaced by $\text{FNR}(\theta, \varphi)$ if the infimum is taken only over $\varphi \in \mathcal{S}_B$.*

In the Subbotin case with parameter $\zeta > 1$, under the assumption of polynomial sparsity (12), the condition $B_n^2 \leq \frac{1}{3}(n/s_n)\overline{F}_0(a_n^* - \delta_n)$ is satisfied if $B_n^2 \leq \exp((\log n)^\nu)$ for some $\nu \in (0, 1 - 1/\zeta)$; see Remark S-2 in [2].

Let us finally mention that this result entails the sub-optimality of any sparsity preserving procedure with FDR asymptotically above some $\alpha > 0$. That is, Corollary 2 also extends to the more general framework of this section.

3.3. Testability. While Theorem 4 is stated as a minimax result over the set Θ , it is possible to formulate results for a given collection of signal strengths only, in a “pointwise in θ ” sense. This completely solves the question of *characterising* the difficulty of the testing problem for an arbitrary sparse vector θ . We postpone the rigorous statements to the Supplementary Material [2], Section S-4; see Theorem S-1 therein. Given its importance we formulate a corollary thereof.

COROLLARY 3 (Difficulty of testing for arbitrary sparse θ). *Fix $\alpha \in (0, 1)$. Suppose we know $\theta \in \ell_0[s_n]$ has s_n nonzero coordinates with arbitrary absolute values $|\theta_j|$, $j \in S_\theta$, whose values may be known or unknown but whose positions and signs are unknown. Then there exists a multiple testing procedure with \mathfrak{R} -risk asymptotically less than α for this problem if and only if, for a_n^* as in Assumption 1, and under 1A,*

$$\overline{\lim}_n \frac{1}{s_n} \sum_{j \in S_\theta} F_{|\theta_j|}(a_n^*) \leq \alpha,$$

and respectively, smaller than or equal to $(1 + \alpha)/2$ under 1B.

Hence, given different arbitrary shapes of signals (again as in Figure 1), to compare the difficulty of the corresponding multiple testing problems, it suffices to compare the corresponding limsups in the last display: the value of the limsup characterises the (asymptotic) difficulty of the multiple testing problem for any *arbitrary* θ with s_n nonzero coordinates.

4. Illustrations. In this section, we present numerical illustrations for the results stated above, as well as provide some intuition behind these.

4.1. Illustrating FDR/FNR tradeoff. Theorems 3 and 6 establish that the FNR is asymptotically the driving force for optimal procedures: any optimal (sparsity preserving) procedure has an FDR vanishing when n goes to infinity and only the FNR matters in the minimax risk.

This phenomenon is illustrated in Figure 2 for thresholding based procedures $\varphi_t = \mathbf{1}_{|X_i| \geq t}$, $t \in \mathbb{R}$. For ease of computations, instead of the FDR, we consider here the closely related marginal FDR (corresponding to the ratio of expectations in (5)), denoted by mFDR. Hence, in the Subbotin setting of Example 2, we consider the marginal risk

$$(26) \quad m\mathfrak{R}(\varphi_t) = \text{mFDR}(\varphi_t) + \text{FNR}(\varphi_t),$$

$$(27) \quad \text{FNR}(\varphi_t) = \overline{\Phi}_\zeta(t - a_n^*) + \overline{\Phi}_\zeta(t + a_n^*),$$

$$(28) \quad \text{mFDR}(\varphi_t) = \frac{2(n - s_n)\overline{\Phi}_\zeta(t)}{2(n - s_n)\overline{\Phi}_\zeta(t) + s_n(1 - \text{FNR}(\varphi_t))},$$

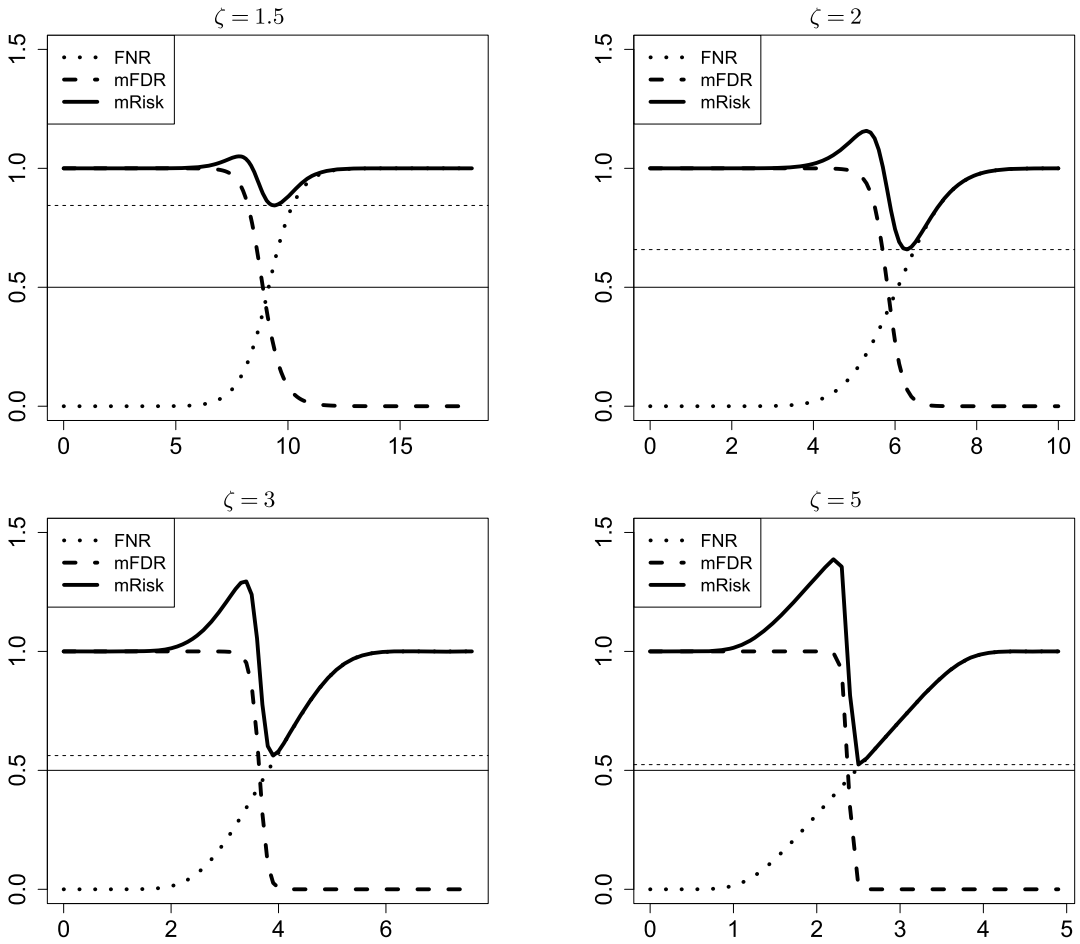


FIG. 2. Finite-sample mFDR (28), FNR (27) and m \mathfrak{R} (26) for a thresholding procedure $\varphi_t = \mathbf{1}_{|X_i| \geq t}$, as a function of the threshold t (x -axis). Here, $n = 10^{10}$, $s_n = 10^2$. The solid horizontal line marks the asymptotic minimax risk of $1/2$, and is approached by the dashed horizontal denoting $m\mathfrak{R}(\varphi)$ for φ an optimal thresholding procedure.

for which the parameter θ is chosen on the border of $\Theta_b = \Theta(a_b, s_n)$ (for $b = 0$) as follows: $\theta_i = a_n^* = \{\zeta \log n/s_n\}^{1/\zeta}$ if $1 \leq i \leq s_n$ and $\theta_i = 0$ otherwise. (It can be shown that our main theorems also hold with the marginal risk $m\mathfrak{R}$ instead of the original risk \mathfrak{R} .)

We can make the following general comments: first, as the threshold t increases (along the x -axes in Figure 2), fewer rejections are made, so that $FNR(\varphi_t)$ increases with t and $mFDR(\varphi_t)$ decreases, hence there is a tradeoff for the finite sample (marginal) risk $m\mathfrak{R}(\varphi_t)$: the minimum is displayed with the light dashed horizontal line. This finite sample risk can be compared to its asymptotic counterpart (thin solid horizontal line). We note that the asymptotic regime is not reached for the current choice of n, s_n if $\zeta = 1.5$, but is approached as ζ increases, with almost a perfect matching when $\zeta = 5$. This is well expected from the convergences rates, for example, obtainable in Theorem 5, that are faster for ζ larger so that the case $\zeta = 5$ well approximates the asymptotic picture. Second, when the asymptotics are (close to being) reached, we see that the dominating part in the risk at the point where the risk is minimum is the FNR, as Theorems 3 and 6 establish. The pictures above give an interpretation of these results: since the transition from 0 to 1 is much more abrupt for the (m)FDR, we should make the (m)FDR close to 0 to make a good tradeoff; the mFDR is close to a step function,

so achieving $mFDR < 1 - \varepsilon$ requires essentially the same threshold as achieving $mFDR < \varepsilon$, hence to make the optimal tradeoff will require $mFDR$ close to zero.

This sharp transition can be seen theoretically from Assumption 1 (see also Remark 3); indeed it follows from (27)–(28) that the $mFDR$ of the φ_t is, with $\overline{F}_0 = \overline{\Phi}_\zeta$,

$$\frac{1}{1 + [1 - \overline{F}_0(t - a_n^*) + o(1)][2((n - s_n)/s_n)\overline{F}_0(t)]^{-1}}.$$

By (16), this tends to zero for $t = a_n^*$, and by (17), it tends to 1 for $t = a_n^* - \delta_n$.

4.2. *Illustrating minimax risks for two signal strengths.* The asymptotic minimax risk found in Section 3 is a function of the multiple signal strengths that we propose to illustrate in this section. For simplicity, we focus on the case of the Subbotin location model with two signal strengths; see Example 5, which corresponds to considering the functional

(29) $\Lambda_\infty : (x, y) \in \mathbb{R}^2 \mapsto \Lambda_\infty(x, y) = \beta \overline{\Phi}_\zeta(\max(x, y)) + (1 - \beta) \overline{\Phi}_\zeta(\min(x, y)).$

Level sets of this function are displayed in Figure 3 for $\beta \in \{1/4, 1/2\}$ and $\zeta \in \{2, 4\}$. As we can see, while the behaviour on the diagonal $x = y$ matches that of the single signal strength case, the asymptotic minimax risk has various behaviours off this diagonal, where there are two different signal strengths. For instance, when $\beta = 1/2$, we see that the signal

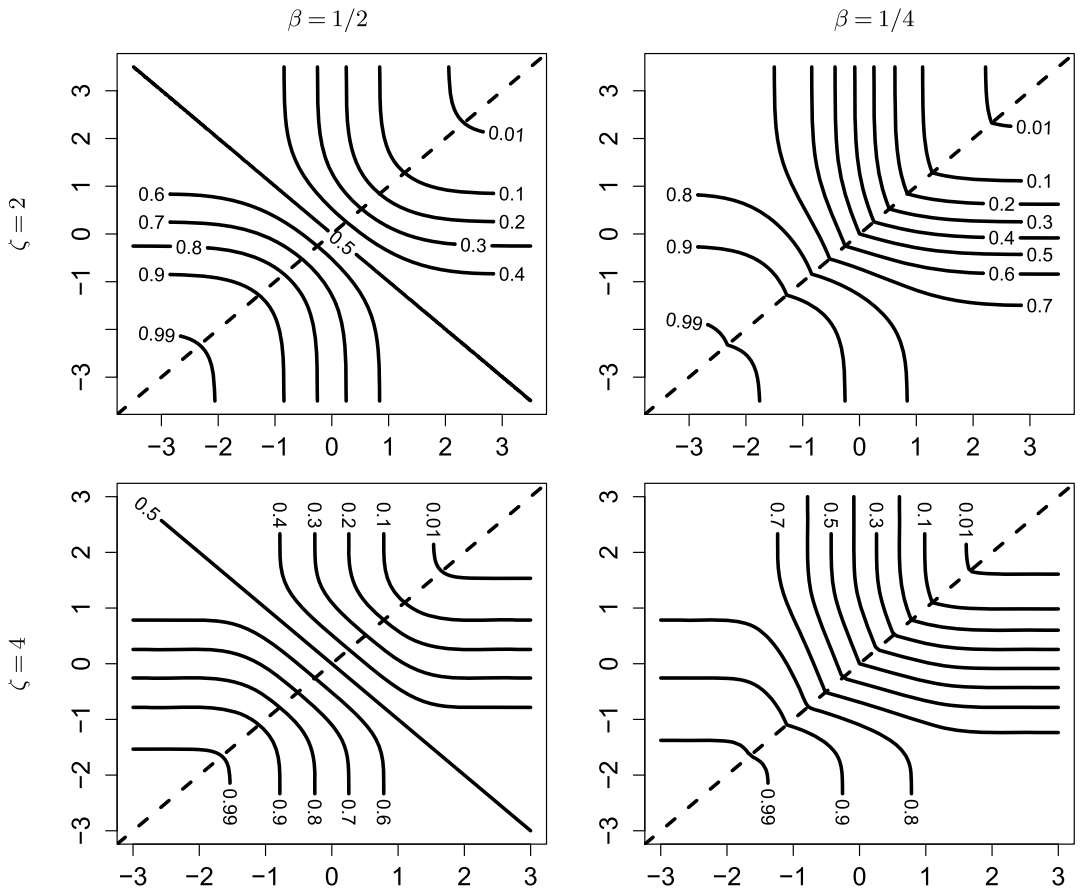


FIG. 3. Level sets of the asymptotic boundary, that is, of the function $\Lambda_\infty(x, y)$ as defined by (29) in the case of two signal strengths, with $\lfloor s_n \beta \rfloor$ nonzero means equal to $\{\zeta \log n / s_n\}^{1/\zeta} + \max(x, y)$ and $s_n - \lfloor s_n \beta \rfloor$ nonzero means equal to $\{\zeta \log n / s_n\}^{1/\zeta} + \min(x, y)$. Left: $\beta = 1/2$; right: $\beta = 1/4$. Top: Gaussian noise; Bottom: Subbotin noise with $\zeta = 4$.

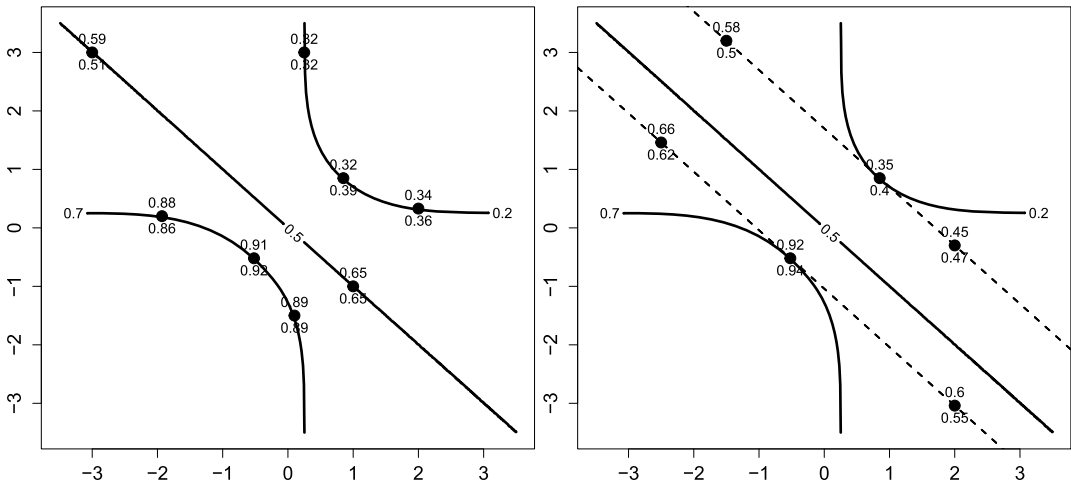


FIG. 4. Same as Figure 3 (top-left, Gaussian case with $\beta = 1/2$) with finite sample risk ($n = 10^6$, $s_n = 20$) at some particular configurations (displayed by black dots). The risk is computed via 100 Monte-Carlo simulations. Two (asymptotically) minimax procedures are implemented: the ℓ -value procedure ($t = 0.3$, risk displayed below each dot) and the BH procedure ($\alpha = 0.1$, risk displayed above each dot). Left: dots are located on the Λ_∞ -level sets of values in $\{0.7, 0.5, 0.2\}$ of the boundary function. Right: dots are located on lines such that the average of x and y is kept constant (not Λ_∞ -level sets).

strengths $(x, -x)$ and $(0, 0)$ are equally risky ($\Lambda_\infty(0, 0) = \Lambda_\infty(x, -x)$ for $x > 0$) while $(1, 3)$ is more risky than $(2, 2)$ (this can be shown by explicit computation or convexity). In general, the shape of the level sets depends on the proportion β and on ζ . First, β strongly affects the level sets: Λ_∞ is larger for $\beta = 1/4$ than for $\beta = 1/2$ off the diagonal $x = y$. This is to be expected because when $\beta = 1/4$, only a proportion $1/4$ of the nonzero means are equal to the larger value $\{\zeta \log n/s_n\}^{1/\zeta} + \max(x, y)$ while the other nonzero values are equal to $\{\zeta \log n/s_n\}^{1/\zeta} + \min(x, y)$. This is clearly a less favorable situation compared to the case where the proportions are balanced ($\beta = 1/2$). Second, ζ also affects the level sets (although less severely): increasing ζ makes the level sets flatter in the center of the (x, y) -picture. This difference comes from the fact that the tails of the ζ -Subbotin distribution are lighter for larger values of ζ .

Figure 4 further illustrates how the (asymptotic) level sets of Λ_∞ are approached for large finite samples: it displays some finite-sample risks $\mathfrak{R}(\theta, \varphi)$ ($n = 10^6$, $s_n = 20$) of the minimax procedures described in Section 1.6: ℓ -value for $t = 0.3$ and BH procedure at level $\alpha = 0.1$, for some parameter θ corresponding to configurations (x, y) taken from the Gaussian and $\beta = 1/2$ plot (top-left panel of Figure 3; see again Example 5). These pointwise risks are estimated via 100 Monte-Carlo simulations. As expected from the slow convergence already discussed above, the finite sample risk has not yet converged to the minimax risk. However, we can see that the global monotonicity of the risk is maintained. More precisely, the left panel of Figure 4 displays the finite sample risks at points taken along asymptotic level sets. We can see that the values reported for the finite sample risks are near-constant along asymptotic level sets (up to the Monte-Carlo errors), suggesting that the shapes of the finite sample level sets are close to those of their asymptotic counterparts. One exception for which the convergence appears to be slower is for configuration $(x, y) = (-3, 3)$: nonasymptotically this configuration seems to be easier than, for example, $(x, y) = (1, -1)$ (although asymptotically equivalent). One possible explanation is that an extreme value of y makes it easily detectable for finite n . On the other hand, the right panel of Figure 4 displays the finite sample risks for points taken along lines which are not asymptotic level sets. The results are markedly different from the left panel: the values vary much more along these lines. Hence, these simple

lines, which are not asymptotic level sets, are also far from being finite-sample levels sets. This suggests that the functional Λ_∞ can be used for comparing testing difficulties of given collections of signal strengths for n large.

5. Extensions to other risks: FDR-controlling procedures and classification. Results in this section are for simplicity stated in the Gaussian location model under beta-min conditions of Section 2, but versions also exist in the general setting of Section 3.

5.1. *Combined testing risk with given FDR control.* The following proposition is easily obtained from our results, but particularly relevant for multiple testing since common intuition says that FDR can be traded for FNR and vice-versa. From this perspective, an interesting alternative risk to \mathfrak{R} is, with $z_+ = z \vee 0 = \max(z, 0)$,

$$\mathfrak{R}_\alpha(\theta, \varphi) = (\text{FDR}(\theta, \varphi) - \alpha)_+ + \text{FNR}(\theta, \varphi).$$

Proposition 1 shows that allowing for some false discoveries by targeting $\mathfrak{R}_\alpha \rightarrow 0$ rather than $\mathfrak{R} = \mathfrak{R}_0 \rightarrow 0$ does not allow for weaker boundary conditions.

PROPOSITION 1. *For $\alpha \in [0, 1)$ and a fixed real b , consider the set of sparsity preserving procedures $\mathcal{S}_B(\Theta_b)$ as in Definition 1. Under the conditions of Theorem 3,*

$$\inf_{\varphi \in \mathcal{S}_B(\Theta_b)} \sup_{\theta \in \Theta_b} \mathfrak{R}_\alpha(\theta, \varphi) = \overline{\Phi}(b) + o(1).$$

Proposition 1 is an immediate consequence of Theorems 1 and 3 and of the bounds $\text{FNR} \leq \mathfrak{R}_\alpha \leq \mathfrak{R}$. From the multiple testing literature perspective, this result is relatively counter-intuitive as allowing a looser FDR control is generally considered to be beneficial for the power of the procedure. This is of course true, but as was shown in Figure 2 and the discussion thereof, for large n the FDR curve is much steeper than the FNR curve, so that the two cannot be traded efficiently (at least for the noise distributions considered here).

5.2. *Classification: Sharp adaptive minimaxity.* The classification (Hamming) loss in terms of classes $\{0\}$ and $\mathbb{R} \setminus \{0\}$ is defined by

$$(30) \quad L_C(\theta, \varphi) = \sum_{i=1}^n (\mathbf{1}\{\theta_i = 0\} \mathbf{1}\{\varphi_i \neq 0\} + \mathbf{1}\{\theta_i \neq 0\} \mathbf{1}\{\varphi_i = 0\}).$$

This risk is studied for sparse vectors in [14] (see also [13]), with a focus on different regimes of large signals. A testing procedure $\varphi \in \{0, 1\}^n$ is said to *achieve almost full recovery* [14] with respect to the Hamming loss over a subset $\Theta_{s_n} \subset \ell_0[s_n]$ if, as $n \rightarrow \infty$,

$$\sup_{\theta \in \Theta_{s_n}} E_\theta L_C(\theta, \varphi) / s_n = o(1).$$

This notion is quite close (although not equivalent) to the notion of conservative testing as considered in Corollary 1; see Section S-3 in [2] for a precise comparison. A complementary notion from [14] is that of *exact recovery*, which will be relevant in Section 6 below.

To allow for a direct comparison with [14], we introduce the class considered therein, defined, for $\Theta(a_b, s)$ as in (10), by

$$(31) \quad \Theta'_b = \Theta'_b(s_n) = \bigcup_{0 \leq s \leq s_n} \Theta(a_b, s).$$

Note that all our lower bound results imply the same bounds for this larger class $\Theta'_b \supset \Theta_b = \Theta(a_b, s_n)$, and the following result gives a corresponding upper bound on this class. We restrict to Gaussian noise to fix ideas, but as before, similar results hold more generally.

THEOREM 7. *Consider the Gaussian location model of Example 1. If one sets $a_b = \sqrt{2 \log(n/s_n)} + b$ for an arbitrary real b or a sequence $b = b_n \rightarrow \pm\infty$, then for $\Theta'_b = \Theta'_b(s_n)$ as in (31), the sharp asymptotic minimax risk for classification is*

$$\inf_{\varphi} \sup_{\theta \in \Theta'_b} E_{\theta} L_C(\theta, \varphi)/s_n = \overline{\Phi}(b) + o(1).$$

For $\varphi = \varphi^{\hat{\ell}}$ the empirical Bayes ℓ -value procedure (S-29) in [2] or $\varphi = \varphi^{\text{BH}}$ the BH procedure (S-18) in [2] at a level $\alpha = \alpha_n = o(1)$ with $-\log \alpha = o((\log n)^{1/2})$ (additionally assuming the polynomial sparsity (12)), the bound is achieved: for any real b , or for $b = b_n \rightarrow \pm\infty$,

$$\sup_{\theta \in \Theta'_b} E_{\theta} L_C(\theta, \varphi)/s_n = \overline{\Phi}(b) + o(1).$$

Theorem 7 is proved in Section S-6.3 in [2]. These results complement some of the results of Sections 4–5 in [14] in the almost sure recovery regime. Theorem 5.2 therein states that there exists an adaptive procedure that achieves almost full recovery if $b_n \gtrsim \log \log n$. Theorem 7 shows that the ℓ -value procedure achieves it under the weakest possible condition $b_n \rightarrow +\infty$, not requiring a particular rate. Further, Theorem 7 investigates the setting of a finite b . Over the class Θ'_b , Theorem 4.2(ii) in [14] provides an in-expectation lower bound that is asymptotically similar to that from Theorem 7, but only for the case of $b \geq 0$ (which essentially amounts to $W > 0$ in the notation from [14]); Theorem 7 provides the sharp asymptotic constant for any real b , and asserts that it can be achieved by an adaptive (i.e., not dependent on s_n or b) procedure.

In-probability results complementing Theorem 7, covering also the case of general signals, are provided in Section S-12 in [2]. Let us remark that in the above result, in contrast to Theorem 2 (and Remark 1), we are able to include the case $b = b_n \rightarrow -\infty$ for the ℓ -value procedure without extra assumptions.

6. Large signal regime and faster minimax rates. In this section, we work in the Subbotin location model presented in Example 2 (sometimes restricted to the Gaussian case for simplicity). An anonymous referee suggested to investigate a “large signal” regime for which nonzero signals have an amplitude larger than $M = M(r) = (\zeta r \log n)^{1/\zeta}$ while the sparsity satisfies, say, $s_n \asymp n^{1-\beta}$, for some $r > \beta$. In Section 6.1, we identify the minimax risk in this large signal regime. In Section 6.2, we consider the possibility of adaptation to both the unknown sparsity and signal strength.

Let us first formally introduce an appropriate parameter space of large signals. For some $\beta \in (0, 1)$, $\zeta > 1$, and $r > \beta$, for fixed $0 < a < b$, define

$$(32) \quad \Theta(r, \beta) = \bigcup_{s_n \in [an^{1-\beta}, bn^{1-\beta}]} \{ \theta \in \ell_0[s_n] : |S_{\theta}| = s_n, |\theta_i| \geq M(r) \text{ for all } i \in S_{\theta} \},$$

$$(33) \quad M = M(r) = (\zeta r \log n)^{1/\zeta}.$$

A main motivation for considering classes (32) is to investigate the rate at which the maximum risk goes to 0 in the second part of Corollary 1 (i.e., Theorem 1 for $b \rightarrow +\infty$ fast). The recent work [37] considers a related problem for the classification loss, but without looking at precise convergence rates. Let us mention that all our results below are also valid for the (normalised) classification loss.

6.1. *Fast minimax rate.* The following result holds and resembles Theorem 2 of [41] for Subbotin noise. There are important differences: first, the latter work studied so-called “generalised Gaussian” noises which do not contain the Gaussian or Subbotin noises considered here, and the obtained rates then differ by logarithmic factors; second, Theorem 8 below provides the minimax rate over *all* possible estimators, not only thresholding estimators; finally, the nonzero signals in class (32) are not necessarily equal (if they were be, averaging procedures could be considered that would allow one to estimate signal strength more easily in some cases).

THEOREM 8. *Consider the Subbotin location model (Example 2) for some $\zeta > 1$ and the parameter set $\Theta(r, \beta)$ defined by (32)–(33). Let $\kappa = \kappa(r, \beta, \zeta)$ be the unique element of $(0, r/2^\zeta)$ such that $(r^{1/\zeta} - \kappa^{1/\zeta})^\zeta - \kappa = \beta$. Then the thresholding procedure $\varphi_i^* = \mathbf{1}_{|X_i| \geq t_n^*}$ based upon the threshold*

$$(34) \quad t_n^* = (\zeta \log n)^{1/\zeta} (r^{1/\zeta} - \kappa^{1/\zeta}) = (\zeta(\beta + \kappa) \log n)^{1/\zeta}$$

is asymptotically rate minimax:

$$\sup_{\theta \in \Theta(r, \beta)} \mathfrak{R}(\theta, \varphi^*) \asymp \inf_{\varphi} \sup_{\theta \in \Theta(r, \beta)} \mathfrak{R}(\theta, \varphi) \asymp n^{-\kappa} / (\log n)^{1-1/\zeta}.$$

The same result holds for classification upon replacing \mathfrak{R} by $E_\theta L_C / n^{1-\beta}$.

Theorem 8 is proved in Section S-9.5. Let us first notice that the threshold t_n^* in (34) is markedly larger than the threshold $(\zeta \beta \log n)^{1/\zeta}$ optimal in the boundary case (see Theorem 4). In particular, if r is much larger than β , we have $\kappa \approx r/2^\zeta$ and $t_n^* \approx 0.5(\zeta r \log n)^{1/\zeta}$, and the minimax risk is converging to 0 roughly at the rate $n^{-r/2^\zeta}$. Second, an important observation is that the FNR and FDR are of the same order in this regime (see the upper bound in the proof). In particular, the FNR part is not dominating for the class of signals (32), which is markedly different from the results we obtained in the regime where the signal is at the boundary (see Theorems 3 and 6, and Theorem S-2 in [2]).

For the classification risk $E_\theta L_C$, we note that the bound achieved using Theorem 8 goes to 0 if $\kappa \geq 1 - \beta$, in which case one has exact recovery in the sense of [14]. We also note that using similar arguments, we can obtain minimax optimal rates also in probability for L_C ; also, for large signals, one can show that the support of θ is recovered with probability going to 1 and study rates for the probability of making at least one error; see Theorem S-12 in [2].

6.2. *Adaptation to fast minimax rates.* We now study possible adaptation in the context of the minimax result of Theorem 8. The class $\Theta(r, \beta)$ depends on two parameters and so adaptation can be considered with respect to both. We focus on the Gaussian case $\zeta = 2$ for simplicity: this case already captures the main phenomena at stake (the first two cases of Theorem 9 below will in fact be obtained also for Subbotin noise). The “top- s_n procedure” will refer to a procedure that selects exactly the s_n largest observations in absolute value. The following theorem summarises our results on the possibility of adaptation. Precise statements can be found in Section S-9 in [2].

THEOREM 9 (Adaptation to large signals, summary in Gaussian case). *Consider the setting of Theorem 8 with $\zeta = 2$ (Gaussian noise). The following points consider the adaptation problem over the class $\Theta(r, \beta)$ for some $\beta \in (0, 1)$ and $\beta < r$. The results hold for the \mathfrak{R} -risk as well as for the normalised classification risk $E_\theta L_C(\theta, \varphi) / n^{1-\beta}$.*

- (i) *Simultaneous adaptation to (r, β) is impossible over the full range of parameters.*
- (ii) *Known exact sparsity s_n , unknown r : the top- s_n procedure provides adaptation to r .*
- (iii) *Unknown β , known r : a plug-in procedure provides adaptation to β .*
- (iv) *Simultaneous adaptation to (r, β) is possible in the regime $\kappa(r, \beta, 2) < 1 - \beta$.*

Moreover, for point (i) the loss due to adaptation is polynomial in n .

Theorem 9 can be interpreted as follows: in the case of large signals in the class (32), when r becomes too large, the rate goes very quickly to zero and adaptation becomes impossible. The intuitive reason is that for very large signals even only one error in the recovery of the support can make the rate drop. In fact, the region in which adaptation is not possible is the one where the optimal (nonnormalised) classification risk is $o(1)$, which corresponds to the region of exact recovery considered in [14].

The results also provide a simple nonartificial example of a setting where there is a polynomial-in- n loss for adaptation (as opposed to more commonly observed logarithmic-in- n losses in nonparametrics).

REMARK 5 (Suboptimality of the top- s_n procedure). The top- s_n procedure, which is adaptive to r here for large signals, is by contrast not sharp minimax (even though it is oracle) when less signal strength is present: the no trade-off result Theorem 3 actually enables one to derive its precise risk for the class Θ_b ; see Corollary S-3 in [2].

7. Discussion. Overview of the results.

This work derived new results for multiple testing from a minimax point of view, in particular deriving the sharp minimax constant for the sum risk $\mathfrak{R} = \text{FDR} + \text{FNR}$. Allowing for a variety of possible noise distributions including standard Gaussian noise, we first considered the beta-min condition on signals, and next allowed for arbitrary sparse signals. This enables one to qualitatively compare difficulties of multiple testing problems, such as the ones depicted on Figure 1. For such general signals, a notable finding is that the overall testing difficulty can be expressed in terms of a necessary and sufficient condition on a certain average, Λ_n in (22), where the strength of a signal is formulated by comparison to the “oracle threshold” (e.g., $\sqrt{2 \log(n/s_n)}$ for Gaussian noise).

As a special case, it follows from this work that the “boundary” for the testing problem corresponds to slightly weaker signals (i.e., classes Θ_b with fixed $b = \overline{\Phi}^{-1}(t)$ for a target risk t in the Gaussian case) compared to classification with almost sure support recovery, for which one needs to be well above the oracle threshold ($b = b_n \rightarrow +\infty$).

The reason is that, as is common in (multiple) testing, one allows for a certain percentage of error in the overall testing risk. An important message regarding this tolerance level is that asymptotically the two types of errors, FDR and FNR, are not symmetric in terms of their contribution to this level: for any optimal procedure, as long as it is sparsity preserving, the FNR dominates, at least for the noise distributions considered here. This main finding of the paper has consequences for (sub-)optimality of some popular procedures, and also intuitively explains why we are able to obtain similar results for multiple testing and classification losses, even at level of sharp constants.

It can be noted that FDR and FNR themselves are by definition not completely symmetric, as the normalisation is by the number of discoveries and the sparsity, respectively. However, the multiple testing problem itself is not symmetric, as the null hypotheses are more numerous by the sparsity assumption. Also, our results cover the (normalised by s_n -) classification loss as well, which itself is symmetric in terms of the two errors. One could also consider other forms of type I errors, beyond FDR or normalised type I-Hamming risk. Our theory then implies that if one considers the FNR (which by its definition equals the normalised

type II-Hamming risk), at least under the fairly natural sparsity-preserving assumption, the contribution of that type II error is the one given above for boundary signals. Finally, note that the dissymmetry in the contribution of FDR and FNR is not present any more for very large signals as investigated in Section 6.

Although originally one main goal was to investigate the phase transition in terms of constants from “easy” to “impossible” multiple testing, the techniques developed are useful for the case of strong signals too. In this setting, very different results are derived: while adaptation to the sharp constant is possible for the worst-case risk, full adaptation to the rate of decrease of the multiple testing risk is impossible for large signals, despite the setting being “easier.” Adaptation becomes possible when restricting to certain subregions of signal/sparsity or assuming one of the two is known.

Adaptive procedures: comparing BH and ℓ -values. We have shown that two popular procedures reach the optimal (local minimax) bounds in an adaptive manner: the ℓ -value procedure and a properly tuned BH procedure. Both have optimal behaviour under relatively similar conditions, but each has its own merits in certain settings. One important message is that, even if the overall risk \mathfrak{R} is allowed to be at least α , taking the standard BH procedure with fixed α parameter does not lead to an optimal risk: taking $\alpha = \alpha_n$ to go to zero is really needed in order to achieve optimality (see Corollary 2).

The ℓ -value procedure does not need a tuning parameter going to zero: it can be applied, for example, for $t = 1/2$. One can interpret this as the fact that it is already on the correct “scale” for the \mathfrak{R} -risk (this is perhaps unsurprising, as this procedure relates to the Bayes classifier for the Hamming risk when the prior is correct), while BH is scaled for the FDR, that is, it essentially prescribes the FDR value. Since the FDR has to be negligible for certain parameters for optimal procedures with respect to \mathfrak{R} (for sparsity preserving ones, as directly follows from combining Theorems 4 and 6), this explains why one needs to take $\alpha = \alpha_n \rightarrow 0$ to achieve optimality for BH.

On the other hand, since ℓ -values are in principle not immediately designed for a (frequentist) FDR-control, proving that the ℓ -value procedure controls the FDR can be nontrivial; in [19] it was shown to be the case for arbitrary sparse signals, but required significant technical work (a result that is invoked here only to handle the somewhat degenerate case $b_n \rightarrow -\infty$ in Remark 1).

From a more practical point of view, choosing between these two competitors depends on the pursued aim: a user interested solely in the combined risk may use the ℓ -value procedure because it is more intrinsic and asymptotically optimal for a fixed value of the parameter t (in the case of the BH- α procedure the level α parameter needs to be tuned appropriately), while if the FDR value should be also controlled or known, they could use the BH procedure with a level α satisfying the convergence requirements. Also note that we have been able to prove optimality of the BH procedure under broader noise assumptions than for the ℓ -value procedure, though we conjecture the latter will work in all models considered here.

Future directions. This work paves the way for several future investigations. First, we expect many ideas relevant here for the Gaussian sequence model to transport to multiple testing for more complex models, such as high dimensional linear regression. Second, since the combined risk involves two parts with markedly different behaviours at the boundary, it could be valuable to find another notion of risk making a better balance between these two terms. Deriving such a risk notion would have interesting consequences for building semisupervised machine learning algorithms achieving an appropriate FDR/FNR tradeoff. Finally, the multiple signal framework introduced here is worth investigating for estimation-type risks, where the constant $\Lambda_\infty = \lim_n \Lambda_n$ is expected to still play a key role.

8. Proof of Theorems 1 and 4. In this section, we give the proof of Theorems 1 and 4. In fact, we shall restrict our attention to proving Theorem 4, because Theorem 1 is a direct consequence of it. Indeed, Lemma S-1 in [2] verifies the conditions of Theorem 4 (first bullet point) in the setting of Theorem 1 for $f_a(x) = \phi(x - a)$ and $a_n^* = \sqrt{2 \log(n/s_n)}$; note that for $\mathbf{a} = (a_n^* + b, \dots, a_n^* + b)$ we have $\Theta(\mathbf{a}, s_n) = \Theta_b$ and $\Lambda_n(\mathbf{a}) = \overline{\Phi}(b)$. (In the cases $b \in \mathbb{R}$ fixed or $b = b_n \rightarrow +\infty$, we apply Theorem 4 for n large enough that $b > -(2 \log(n/s_n))^{1/2}$; in the case $b = b_n \rightarrow -\infty$ we replace b_n by a sequence $b'_n \rightarrow -\infty$ satisfying $b'_n > \max(-(2 \log(n/s_n))^{1/2}, b_n)$ before applying Theorem 4, obtaining in this way $\inf_{\varphi} \sup_{\theta \in \Theta_{b'_n}} \mathfrak{R}(\theta, \varphi) = \overline{\Phi}(b'_n) + o(1) = 1 + o(1)$. Since $\Theta_{b'_n} \subset \Theta_{b_n}$ we deduce $\inf_{\varphi} \sup_{\theta \in \Theta_{b_n}} \mathfrak{R}(\theta, \varphi) \geq 1 + o(1)$; since the upper bound in this case is obtained using the trivial test $\varphi \equiv 0$ we deduce Theorem 1 in all cases.)

8.1. *Lower bound.* Let us start by proving the lower bound. For this, we apply the general lower bound of Theorem S-3 in [2] in our particular model: for any $\rho > 0$, any $s_n \geq 1$, any $\eta \in (0, 1)$ and any prior π on \mathbb{R}^n , with P_{π} denoting the joint law of (X, θ) under π ,

$$\inf_{\varphi} \sup_{\theta \in \Theta(\mathbf{a}, s_n)} \mathfrak{R}(\theta, \varphi) \geq \left(\lambda \wedge \frac{\rho \lambda}{1 + \rho \lambda} \right) (1 - e^{-c \eta^2 M_{\rho}}) - n(1 \vee \rho) P_{\pi}(\|\theta\|_0 > s_n) - 2 P_{\pi}(\theta \notin \Theta(\mathbf{a}, s_n)),$$

for some universal constant $c > 0$, where $\lambda = M_{\rho}(1 - \eta)/s_n$ and

$$M_{\rho} = \sum_{i=1}^n P_{\pi}[\theta_i \neq 0, \ell_i(X) > \rho/(1 + \rho)],$$

for $\ell_i(X) = P_{\pi}(\theta_i = 0 | X)$, $1 \leq i \leq n$. We will apply this with $\rho = \rho_n$ and $\eta = \eta_n$ particular positive sequences converging slowly to infinity and 0 respectively, to be specified later on.

Let us consider a specific prior π as a product prior over s_n blocks of consecutive coordinates $Q_1 = \{1, 2, \dots, q\}$, $Q_2 = \{q + 1, \dots, 2q\}, \dots, Q_{s_n} = \{(s_n - 1)q + 1, \dots, n'\}$, where $q = \lfloor n/s_n \rfloor$ and $n' = qs_n$. We write Q_{∞} for the (possibly empty) set $\{n' + 1, \dots, n\}$. Over each block Q_j , $1 \leq j \leq s_n$, one takes the following prior: first draw an integer I_j from the uniform distribution $\mathcal{U}(Q_j)$ over the block Q_j and next for each $i \in Q_j$ set $\theta_i = a_j$ if $i = I_j$ and $\theta_i = 0$ otherwise. For $i \in Q_{\infty}$, set $\theta_i = 0$. With this prior, we clearly have $P_{\pi}(\|\theta\|_0 > s_n) = 0$ and $P_{\pi}(\theta \notin \Theta(\mathbf{a}, s_n)) = 0$. Moreover, for all $1 \leq j \leq s_n$ and $i \in Q_j$,

$$\begin{aligned} \ell_i(X) &= P_{\pi}(i \neq I_j | X) = 1 - w_i^j(X), \\ w_i^j(X) &= \frac{f_{a_j}(X_i)/f_0(X_i)}{\sum_{k \in Q_j} f_{a_j}(X_k)/f_0(X_k)} = \frac{h(X_i, a_j)}{\sum_{k \in Q_j} h(X_k, a_j)}, \\ h(x, a) &= f_a(x)/f_0(x). \end{aligned}$$

In addition,

$$\begin{aligned} M_{\rho} &= \sum_{i=1}^n P_{\pi}[\theta_i \neq 0, \ell_i(X) > \rho/(1 + \rho)] \\ &= \sum_{j=1}^{s_n} \sum_{i \in Q_j} P_{\pi}[i = I_j, \ell_i(X) > \rho/(1 + \rho)] \\ (35) \quad &= \sum_{j=1}^{s_n} \sum_{i \in Q_j} P_{\pi}\left[i = I_j, (\rho + 1)h(X_i, a_j) < \sum_{k \in Q_j} h(X_k, a_j) \right] \end{aligned}$$

$$\begin{aligned} &\geq \sum_{j=1}^{s_n} \sum_{i \in Q_j} P_\pi [i = I_j, \#\{k \in Q_j \setminus \{i\} : h(X_k, a_j) > h(X_i, a_j)\} \geq \rho] \\ &= \sum_{j=1}^{s_n} P_{X_i \sim f_{a_j}} [\#\{k \in Q_j \setminus \{i\} : h(\varepsilon_k, a_j) > h(X_i, a_j)\} \geq \rho], \end{aligned}$$

where the ε_k 's are i.i.d. $\sim f_0$, and in the last line $i = i_j \in Q_j$ is chosen arbitrarily by symmetry.

Consider the first case, when Assumption 1A holds. We see that $\varepsilon_k > X_i$ implies $h(\varepsilon_k, a_j) > h(X_i, a_j)$, so that the last display can be further lower bounded by

$$\sum_{j=1}^{s_n} P_{X_i \sim f_{a_j}} [\#\{k \in Q_j \setminus \{i\} : \varepsilon_k > X_i\} \geq \rho].$$

By Lemma S-7 in [2], provided ρ satisfies the condition of the lemma, we have

$$P_{X_i \sim f_{a_j}} [\#\{k \in Q_j \setminus \{i\} : \varepsilon_k > X_i\} \geq \rho] = F_{a_j}(a_n^*) + o(1),$$

so that continuing the inequalities,

$$(36) \quad M_\rho / s_n \geq s_n^{-1} \sum_{j=1}^{s_n} F_{a_j}(a_n^*) + o(1) = \Lambda_n(\mathbf{a}) + o(1),$$

where $\Lambda_n(\mathbf{a})$ is defined by (22). This gives the lower bound

$$\begin{aligned} \inf_{\varphi} \sup_{\theta \in \Theta(\mathbf{a}, s_n)} \mathfrak{R}(\theta, \varphi) &\geq \left([\Lambda_n(\mathbf{a}) + o(1)](1 - \eta) \wedge \frac{\rho[\Lambda_n(\mathbf{a}) + o(1)](1 - \eta)}{1 + \rho[\Lambda_n(\mathbf{a}) + o(1)](1 - \eta)} \right) \\ &\quad \times (1 - e^{-c\eta^2[\Lambda_n(\mathbf{a}) + o(1)](1 - \eta)s_n}). \end{aligned}$$

Now using that for all $x \in [0, 1]$, $A, y > 0$, we have

$$[x \wedge (yx / (1 + yx))](1 - e^{-Ax}) \geq x + 0 \wedge (1 - 1 / (1 + yx) - x) - xe^{-Ax} \geq x - 1/y - 1/(Ae),$$

we deduce

$$\inf_{\varphi} \sup_{\theta \in \Theta(\mathbf{a}, s_n)} \mathfrak{R}(\theta, \varphi) \geq [\Lambda_n(\mathbf{a}) + o(1)](1 - \eta) - \rho^{-1} - 1 / (c\eta^2 s_n e).$$

Let us set $\rho = \lfloor (n / (3s_n)) \overline{F}_0(a_n^* - \delta_n) \rfloor \rightarrow \infty$, an admissible choice for applying Lemma S-7 in [2]. Further setting $\eta = s_n^{-1/4}$, one gets

$$\liminf_n \inf_{\varphi} \sup_{\theta \in \Theta(\mathbf{a}, s_n)} \mathfrak{R}(\theta, \varphi) \geq \Lambda_n(\mathbf{a}) + o(1).$$

This proves the lower bound.

When instead Assumption 1B holds, we note that $h(x, a) = h(-x, a)$ is increasing in $|x|$ for $a \neq 0$, hence returning to (35) we see

$$M_\rho \geq \sum_{j=1}^{s_n} P_{X_i \sim f_{a_j}} [\#\{k \in Q_j \setminus \{i\} : |\varepsilon_k| > |X_i|\} \geq \rho].$$

Lemma S-7 in [2] tells us that in this case we have

$$P_{X_i \sim f_{a_j}} [\#\{k \in Q_j \setminus \{i\} : |\varepsilon_k| > |X_i|\} \geq \rho] = 1 - 2\overline{F}_{a_j}(a_n^*) + o(1) = 2F_{a_j}(a_n^*) - 1 + o(1),$$

yielding

$$(37) \quad M_\rho / s_n \geq 2\Lambda_n(\mathbf{a}) - 1 + o(1),$$

and we finish the proof as under the other assumption.

8.2. *Upper bound.* If a subsequence n_j is such that $\Lambda_{n_j} \rightarrow 1$, then the trivial test $\varphi = 0$ has $\mathfrak{R}(\theta, \varphi) = 1 = \Lambda_{n_j} + o(1) = 2\Lambda_{n_j} - 1 + o(1)$. We may therefore limit our attention to a subsequence on which Λ_{n_j} is bounded away from 1. Relabel n_j as n .

Define

$$\varphi_i(X) = \mathbf{1}\{|X_i| > a_n^*\}.$$

Write V and S for the number of false discoveries and true discoveries, respectively, made by φ .

By Markov’s inequality, for any $c_n > 0$ the number of false discoveries satisfies

$$P_\theta(V > c_n s_n) \leq c_n^{-1} s_n^{-1} E[V] = c_n^{-1} (2(n - s_n)/s_n) \bar{F}_0(a_n^*).$$

This latter expression tends to zero by Assumption 1 if c_n tends to zero slowly enough, yielding that $V = o_P(s_n)$. Similarly, using that $\text{Var}(S) \leq E[S] \leq s_n$ and applying Chebyshev’s inequality, the number of true discoveries satisfies

$$P_\theta[|S - E_\theta S| \geq s_n^{3/4}] \leq s_n^{-1/2}.$$

Let \mathcal{A} be an event of probability tending to one on which V and S are suitably bounded.

Under Assumption 1A, for $a \neq 0$ we have by symmetry

$$P_{\theta_i=a}(|X_i| > a_n^*) = \bar{F}_a(a_n^*) + F_a(-a_n^*) = \bar{F}_a(a_n^*) + \bar{F}_{-a}(a_n^*).$$

Noting that by the assumed monotonicity $\bar{F}_{-|a|}(a_n^*) \leq \bar{F}_0(a_n^*) \rightarrow 0$ because $a_n^* \rightarrow \infty$, we deduce that

$$P_{\theta_i=a}(|X_i| > a_n^*) = \bar{F}_{|a|}(a_n^*) + o(1),$$

where the errors $o(1)$ tend to zero uniformly in a . We thus calculate

$$\begin{aligned} E_\theta[S] &= \sum_{i \in S_\theta} P_{\theta_i}(|X_i| > a_n^*) = \sum_{i \in S_\theta} (\bar{F}_{|\theta_i|}(a_n^*) + o(1)) \\ &\geq \sum_{j \leq s_n} (\bar{F}_{a_j}(a_n^*) + o(1)) = s_n[1 - \Lambda_n(\mathbf{a}) + o(1)], \end{aligned}$$

where we have used that S_θ can be enumerated as i_1, \dots, i_{s_n} with $|\theta_{i_j}| \geq a_j > 0$.

The combined risk of φ is then

$$\begin{aligned} \mathfrak{R}(\theta, \varphi) &= E_\theta \frac{V}{V + S} + E_\theta \frac{s_n - S}{s_n} \\ &\leq P(\mathcal{A}^c) + \frac{o(s_n)}{o(s_n) + s_n(1 - \Lambda_n(\mathbf{a}) + o(1)) - s_n^{3/4}} + \frac{s_n(\Lambda_n(\mathbf{a}) + o(1))}{s_n} \\ &= \Lambda_n(\mathbf{a}) + o(1), \end{aligned}$$

yielding the desired upper bound.

Under Assumption 1B, the same calculations hold, except that now $P_{\theta_i}(|X_i| > a_n^*) = 2\bar{F}_{\theta_i}(a_n^*)$, hence (again using the assumed monotonicity) $s_n^{-1} \sum_{i \in S_\theta} P_{\theta_i}(|X_i| > a_n^*) \geq s_n^{-1} \sum_{j \leq s_n} 2\bar{F}_{a_j}(a_n^*) = 1 - (2\Lambda_n(\mathbf{a}) - 1 + o(1))$, which gives the desired upper bound in this case as well.

Acknowledgements. The authors are grateful to Aad van der Vaart for suggesting investigating the case of multiple signals, which led us to the results in Section 3.2. We also thank two anonymous referees and an associate editor for suggesting investigating more general noise models and other comments, which led to substantial generalisations.

Funding. IC acknowledges support from the Institut Universitaire de France. ER has been supported by ANR-16-CE40-0019 (SansSouci), ANR-21-CE23-0035 (ASCAI), the GDR ISIS through the “projets exploratoires” program (project TASTY) and the Emergence project MARS of Sorbonne Université.

ER and IC have been supported by ANR-17-CE40-0001 (BASICS) and ANR-23-CE40-0018-01 (BACKUP).

KA is supported by the EPSRC Programme Grant on the Mathematics of Deep Learning under the project EP/V026259/1. Initial work on this project was completed while he was at Université Paris-Saclay, supported by a public grant as part of the Investissement d’avenir project, reference ANR-11-LABX-0056-LMH, LabEx LMH.

SUPPLEMENTARY MATERIAL

Supplement to: Sharp multiple testing boundary for sparse sequences (DOI: [10.1214/24-AOS2404SUPP](https://doi.org/10.1214/24-AOS2404SUPP); .pdf). The file [2] includes the remaining proofs for the main paper, and some further results and discussions.

REFERENCES

- [1] ABRAHAM, K., CASTILLO, I. and ROQUAIN, É. (2022). Empirical Bayes cumulative ℓ -value multiple testing procedure for sparse sequences. *Electron. J. Stat.* **16** 2033–2081. MR4415394 <https://doi.org/10.1214/22-ejs1979>
- [2] ABRAHAM, K., CASTILLO, I. and ROQUAIN, E. (2024). Supplement to “Sharp multiple testing boundary for sparse sequences.” <https://doi.org/10.1214/24-AOS2404SUPP>
- [3] ABRAMOVICH, F., BENJAMINI, Y., DONOHO, D. L. and JOHNSTONE, I. M. (2006). Adapting to unknown sparsity by controlling the false discovery rate. *Ann. Statist.* **34** 584–653. MR2281879 <https://doi.org/10.1214/009053606000000074>
- [4] ARIAS-CASTRO, E., CANDÈS, E. J. and PLAN, Y. (2011). Global testing under sparse alternatives: ANOVA, multiple comparisons and the higher criticism. *Ann. Statist.* **39** 2533–2556. MR2906877 <https://doi.org/10.1214/11-AOS910>
- [5] ARIAS-CASTRO, E. and CHEN, S. (2017). Distribution-free multiple testing. *Electron. J. Stat.* **11** 1983–2001. MR3651021 <https://doi.org/10.1214/17-EJS1277>
- [6] BANERJEE, S., CASTILLO, I. and GHOSAL, S. (2021). Bayesian inference in high-dimensional models. Book chapter to appear in Springer volume on data science. Preprint. Available at [arXiv:2101.04491](https://arxiv.org/abs/2101.04491).
- [7] BARAUD, Y. (2002). Non-asymptotic minimax rates of testing in signal detection. *Bernoulli* **8** 577–606. MR1935648
- [8] BELITSER, E. and NURUSHEV, N. (2022). Uncertainty quantification for robust variable selection and multiple testing. *Electron. J. Stat.* **16** 5955–5979. MR4515719 <https://doi.org/10.1214/22-ejs2088>
- [9] BENJAMINI, Y. and HOCHBERG, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* **57** 289–300. MR1325392
- [10] BENJAMINI, Y. and YEKUTIELI, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.* **29** 1165–1188. MR1869245 <https://doi.org/10.1214/aos/1013699998>
- [11] BOGDAN, M., CHAKRABARTI, A., FROMMLET, F. and GHOSH, J. K. (2011). Asymptotic Bayes-optimality under sparsity of some multiple testing procedures. *Ann. Statist.* **39** 1551–1579. MR2850212 <https://doi.org/10.1214/10-AOS869>
- [12] BÜHLMANN, P. and VAN DE GEER, S. (2011). *Statistics for High-Dimensional Data: Methods, Theory and Applications*. Springer Series in Statistics. Springer, Heidelberg. MR2807761 <https://doi.org/10.1007/978-3-642-20192-9>
- [13] BUTUCEA, C., MAMMEN, E., NDAOUD, M. and TSYBAKOV, A. B. (2023). Variable selection, monotone likelihood ratio and group sparsity. *Ann. Statist.* **51** 312–333. MR4564858 <https://doi.org/10.1214/22-aos2251>
- [14] BUTUCEA, C., NDAOUD, M., STEPANOVA, N. A. and TSYBAKOV, A. B. (2018). Variable selection with Hamming loss. *Ann. Statist.* **46** 1837–1875. MR3845003 <https://doi.org/10.1214/17-AOS1572>
- [15] BUTUCEA, C. and STEPANOVA, N. (2017). Adaptive variable selection in nonparametric sparse additive models. *Electron. J. Stat.* **11** 2321–2357. MR3656494 <https://doi.org/10.1214/17-EJS1275>
- [16] CAI, T. T. and SUN, W. (2009). Simultaneous testing of grouped hypotheses: Finding needles in multiple haystacks. *J. Amer. Statist. Assoc.* **104** 1467–1481. MR2597000 <https://doi.org/10.1198/jasa.2009.t08415>

- [17] CAI, T. T., SUN, W. and WANG, W. (2019). Covariate-assisted ranking and screening for large-scale two-sample inference. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **81** 187–234. Includes discussions and a rejoinder. MR3928141
- [18] CASTILLO, I. and MISMER, R. (2018). Empirical Bayes analysis of spike and slab posterior distributions. *Electron. J. Stat.* **12** 3953–4001. MR3885271 <https://doi.org/10.1214/18-EJS1494>
- [19] CASTILLO, I. and ROQUAIN, É. (2020). On spike and slab empirical Bayes multiple testing. *Ann. Statist.* **48** 2548–2574. MR4152112 <https://doi.org/10.1214/19-AOS1897>
- [20] CASTILLO, I. and SZABÓ, B. (2020). Spike and slab empirical Bayes sparse credible sets. *Bernoulli* **26** 127–158. MR4036030 <https://doi.org/10.3150/19-BEJ1119>
- [21] DONOHO, D. and JIN, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.* **32** 962–994. MR2065195 <https://doi.org/10.1214/009053604000000265>
- [22] DURAND, G. (2019). Adaptive p -value weighting with power optimality. *Electron. J. Stat.* **13** 3336–3385. MR4010982 <https://doi.org/10.1214/19-ejs1578>
- [23] EFRON, B., TIBSHIRANI, R., STOREY, J. D. and TUSHER, V. (2001). Empirical Bayes analysis of a microarray experiment. *J. Amer. Statist. Assoc.* **96** 1151–1160. MR1946571 <https://doi.org/10.1198/016214501753382129>
- [24] FROMONT, M., LERASLE, M. and REYNAUD-BOURET, P. (2016). Family-wise separation rates for multiple testing. *Ann. Statist.* **44** 2533–2563. MR3576553 <https://doi.org/10.1214/15-AOS1418>
- [25] GENOVESE, C. and WASSERMAN, L. (2002). Operating characteristics and extensions of the false discovery rate procedure. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64** 499–517. MR1924303 <https://doi.org/10.1111/1467-9868.00347>
- [26] GINÉ, E. and NICKL, R. (2016). *Mathematical Foundations of Infinite-Dimensional Statistical Models*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge Univ. Press, New York. MR3588285 <https://doi.org/10.1017/CBO9781107337862>
- [27] HALL, P. and JIN, J. (2008). Properties of higher criticism under strong dependence. *Ann. Statist.* **36** 381–402. MR2387976 <https://doi.org/10.1214/009053607000000767>
- [28] HALL, P. and JIN, J. (2010). Innovated higher criticism for detecting sparse signals in correlated noise. *Ann. Statist.* **38** 1686–1732. MR2662357 <https://doi.org/10.1214/09-AOS764>
- [29] HELLER, R. and ROSSET, S. (2021). Optimal control of false discovery criteria in the two-group model. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **83** 133–155. MR4220987 <https://doi.org/10.1111/rssb.12403>
- [30] IGNATIADIS, N. and HUBER, W. (2021). Covariate powered cross-weighted multiple testing. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **83** 720–751. MR4319999 <https://doi.org/10.1111/rssb.12411>
- [31] INGSTER, Y. I., TSYBAKOV, A. B. and VERZELEN, N. (2010). Detection boundary in sparse regression. *Electron. J. Stat.* **4** 1476–1526. MR2747131 <https://doi.org/10.1214/10-EJS589>
- [32] INGSTER, Y. I. and SUSLINA, I. A. (2003). *Nonparametric Goodness-of-Fit Testing Under Gaussian Models*. Lecture Notes in Statistics **169**. Springer, New York. MR1991446 <https://doi.org/10.1007/978-0-387-21580-8>
- [33] JIN, J. and KE, Z. T. (2016). Rare and weak effects in large-scale inference: Methods and phase diagrams. *Statist. Sinica* **26** 1–34. MR3468343
- [34] JOHNSTONE, I. M. and SILVERMAN, B. W. (2004). Needles and straw in haystacks: Empirical Bayes estimates of possibly sparse sequences. *Ann. Statist.* **32** 1594–1649. MR2089135 <https://doi.org/10.1214/009053604000000030>
- [35] JOHNSTONE, I. M. and SILVERMAN, B. W. (2005). EbayesThresh: R programs for empirical Bayes thresholding. *J. Stat. Softw.* **12**.
- [36] LI, X. and FITHIAN, W. (2020). Optimality of the max test for detecting sparse signals with Gaussian or heavier tail. Preprint. Available at [arXiv:2006.12489](https://arxiv.org/abs/2006.12489).
- [37] MILLER, J. C. and STEPANOVA, N. A. (2023). Adaptive signal recovery with Subbotin noise. *Statist. Probab. Lett.* **196** Paper No. 109791, 6. MR4549624 <https://doi.org/10.1016/j.spl.2023.109791>
- [38] NEUVIAL, P. and ROQUAIN, E. (2012). On false discovery rate thresholding for classification under sparsity. *Ann. Statist.* **40** 2572–2600. MR3097613 <https://doi.org/10.1214/12-AOS1042>
- [39] NICKL, R. and VAN DE GEER, S. (2013). Confidence sets in sparse regression. *Ann. Statist.* **41** 2852–2876. MR3161450 <https://doi.org/10.1214/13-AOS1170>
- [40] RABINOVICH, M., JORDAN, M. I. and WAINWRIGHT, M. J. (2020). Lower bounds in multiple testing: A framework based on derandomized proxies. Arxiv eprint. Available at [arXiv:2005.03725](https://arxiv.org/abs/2005.03725).
- [41] RABINOVICH, M., RAMDAS, A., JORDAN, M. I. and WAINWRIGHT, M. J. (2020). Optimal rates and trade-offs in multiple testing. *Statist. Sinica* **30** 741–762. MR4214160
- [42] ROQUAIN, E. and VAN DE WIEL, M. A. (2009). Optimal weighting for false discovery rate control. *Electron. J. Stat.* **3** 678–711. MR2521216 <https://doi.org/10.1214/09-EJS430>
- [43] SU, W. and CANDÈS, E. (2016). SLOPE is adaptive to unknown sparsity and asymptotically minimax. *Ann. Statist.* **44** 1038–1068. MR3485953 <https://doi.org/10.1214/15-AOS1397>

- [44] SUN, W. and CAI, T. T. (2007). Oracle and adaptive compound decision rules for false discovery rate control. *J. Amer. Statist. Assoc.* **102** 901–912. MR2411657 <https://doi.org/10.1198/016214507000000545>
- [45] SUN, W. and CAI, T. T. (2009). Large-scale multiple testing under dependence. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 393–424. MR2649603 <https://doi.org/10.1111/j.1467-9868.2008.00694.x>