The Annals of Statistics
2016, Vol. 44, No. 6, 2533–2563
DOI: 10.1214/15-AOS1418

FAMILY-WISE SEPARATION RATES FOR MULTIPLE TESTING[1]

BY MAGALIE FROMONT*, MATTHIEU LERASLE† AND PATRICIA REYNAUD-BOURET†

*Université Bretagne Loire, IRMAR (CNRS, UMR 6625)**
Université de Nice Sophia-Antipolis, LJAD (CNRS, UMR 7351)†

Starting from a parallel between some minimax adaptive tests of a single null hypothesis, based on aggregation approaches, and some tests of multiple hypotheses, we propose a new second kind error-related evaluation criterion, as the core of an emergent minimax theory for multiple tests. Aggregation-based tests, proposed for instance by Baraud [*Bernoulli* **8** (2002) 577–606], Baraud, Huet and Laurent [*Ann. Statist.* **31** (2003) 225–251] or Fromont and Laurent [*Ann. Statist.* **34** (2006) 680–720], are justified through their first kind error rate, which is controlled by the prescribed level on the one hand, and through their separation rates over various classes of alternatives, rates which are minimax on the other hand. We show that some of these tests can be viewed as the first steps of classical step-down multiple testing procedures, and accordingly be evaluated from the multiple testing point of view also, through a control of their Family-Wise Error Rate (FWER). Conversely, many multiple testing procedures, from the historical ones of Bonferroni and Holm, to more recent ones like min-p procedures or randomized procedures such as the ones proposed by Romano and Wolf [*J. Amer. Statist. Assoc.* **100** (2005) 94–108], can be investigated from the minimax adaptive testing point of view. To this end, we extend the notion of separation rate to the multiple testing field, by defining the *weak Family-Wise Separation Rate* and its stronger counterpart, the *Family-Wise Separation Rate* (FWSR). As for nonparametric tests of a single null hypothesis, we prove that these new concepts allow an accurate analysis of the second kind error of a multiple testing procedure, leading to clear definitions of minimax and minimax adaptive multiple tests. Some illustrations in classical Gaussian frameworks corroborate several expected results under particular conditions on the tested hypotheses, but also lead to new questions and perspectives.

1. Introduction. Following the Neyman–Pearson principle in single null hypotheses testing problems, the main concern in multiple testing problems is generally to construct procedures controlling a chosen first kind error-related criterion.

Many first kind error-related criteria for multiple tests have been introduced in the statistical literature, generalizing or relaxing the traditional *Family-Wise Error*

Received January 2015; revised November 2015.

[1]Supported in part by the French Agence Nationale de la Recherche (ANR 2011 BS01 010 01 projet Calibration).

MSC2010 subject classifications. 62G10, 62H15, 62C20.

Key words and phrases. Multiple testing, Family-Wise Error Rate, step-down procedure, separation rate, minimax test, adaptive test.

Rate (FWER) defined as the probability of one or more false discoveries (true null hypotheses that are rejected). Thus, the *Per-Family Error Rate* (PFER) suggested by Spjotvoll [34] corresponds to the average number of false discoveries, while the $k-$ FWER introduced by Hommel and Hoffman [17] and further studied by Korn et al. [21], Lehmann and Romano [24], Romano and Shaikh [28] or Romano and Wolf [30, 31], is the probability of k or more false discoveries. Like Genovese and Wasserman [14], many of these authors also focused on the *False Discovery Proportion* (FDP), whose expected value is the very popular *False Discovery Rate* (FDR) introduced by Benjamini and Hochberg [3].

Up to now, however, very few articles deal with the optimality of multiple tests in terms of second kind error. The articles by Lehmann, Romano and Shaffer [25], and by Romano, Shaikh and Wolf [27] both give maximin type optimality results, but each with a different notion of maximin optimality. While Lehmann, Romano and Shaffer [25] consider the minimum probability of rejecting at least one false hypothesis when at least one hypothesis deviates from the truth at a given degree, Romano, Shaikh and Wolf [27] consider the minimum probability of rejecting at least one hypothesis when the hypotheses are not all true simultaneously.

We propose here new second kind error-related criteria to evaluate multiple procedures whose FWER is controlled by a prescribed level α in $(0,1)$. These criteria are inspired by the minimax theory for nonparametric tests of a single null hypothesis. The minimax testing theory was historically introduced by Ingster in his series of papers [19] from a purely asymptotic point of view. This asymptotic theory does not seem to be the most adequate to import in multiple testing frameworks, as the asymptotics there should concern the number of tested hypotheses as well as the sample size, leading to consider how the number of hypotheses grows with respect to the sample size. We therefore turned toward the nonasymptotic theory introduced by Baraud in [1], which is based on the notion of uniform separation rate over a class of alternatives. Considering a single null hypothesis H_0 and a class of alternatives $\mathcal{Q}$, the uniform separation rate of a level α test over $\mathcal{Q}$ with prescribed second kind error rate β in $(0,1)$ is defined as the minimal distance between the underlying distributions in $\mathcal{Q}$ and H_0 which guarantees that the second kind error rate of the test is at most equal to β (a more precise expression is given later on). The test is then said to be minimax over $\mathcal{Q}$ if its uniform separation rate over $\mathcal{Q}$ achieves the lowest possible value, possibly up to a multiplicative constant. Furthermore, it is said to be minimax adaptive over a collection of classes of alternatives if it is minimax or nearly minimax over every class of the collection.

The literature on minimax and minimax adaptive testing is huge, and provides a now well known and convenient framework to study the theoretical performance of nonparametric tests of single null hypotheses. Beyond the founding articles by Ingster [19] and Baraud [1], many others are devoted to the computation of minimax separation rates over various classes of alternatives, and the construction of minimax or minimax adaptive tests in many statistical models. For the present concerns, one can cite, for instance, [2, 9, 11, 20, 35] or [12].

Our purpose here is to provide such a framework in the multiple testing context. Most of minimax adaptive tests of a single null hypothesis H_0 are constructed from the aggregation of a collection of minimax tests for different single null hypotheses, all including H_0. Therefore, we first investigate the parallel that can be drawn between such minimax adaptive tests, and some classical single-step or step-down multiple testing procedures. We prove in particular that some of the minimax adaptive tests proposed in [2, 9, 11] or [12] for instance are closely related to the Bonferroni-type single-step multiple testing procedures, while others correspond to the first step of a min-p procedure, as defined in [6]. Conversely, a multiple test may be associated with an aggregated test of a null hypothesis contained in all the tested hypotheses, test that can be studied using the minimax theory. This parallel motivates the definition of the first criterion we introduce here: the *weak Family-Wise Separation Rate* denoted by wFWSR, and a stronger second criterion: the (strong) *Family-Wise Separation Rate* denoted by FWSR. This second criterion is in fact the key point to lay the foundations of a minimax theory for multiple tests whose FWER is controlled by a prescribed level α. The FWSR and its corresponding benchmark, the *minimax Family-Wise Separation Rate*, presented in this article, are thus new tools to evaluate the second kind error performance of a multiple test. Considering simple multiple testing problems in Gaussian regression frameworks, we prove for instance that in some cases, the FWSR of all the Bonferroni, Holm and min-p procedures are optimal from this minimax point of view, whereas in other cases, the Bonferroni procedure is clearly sub-optimal. Beyond the evaluation of a multiple test itself, the *minimax Family-Wise Separation Rate* can also be viewed as an indicator of the difficulty or complexity of the considered testing problem. In particular, we exhibit general conditions on the considered hypotheses, which guarantee that the *minimax Family-Wise Separation Rate* for multiple tests is lower bounded by the classical minimax Separation Rate for single tests, thus formalizing the intuition that multiple testing is more difficult than single testing. Through our illustrations in Gaussian regression frameworks, we furthermore prove that when these general conditions are not satisfied, the *minimax Family-Wise Separation Rate* for multiple tests may be smaller than the classical minimax Separation Rate for single tests, which may suggest, looking at things superficially, that multiple testing may be easier than single testing in some cases. This apparent counter-intuitive result in fact leads to a deeper analysis of the introduced criteria, and to a further reflection about the basic nature of a multiple testing problem, focusing on its fundamental differences with single testing problems. The emphasis is here placed on the importance attached, in a multiple testing problem, to each individual tested hypotheses, contrary to an aggregation-based single testing problem where only a single null hypothesis contained in all the tested hypotheses has to be taken into account. These considerations convince us that there is still work to do and encourage us to further develop the minimax theory for multiple tests in future works.

This article is organized as follows. Section 2 contains notation and preliminary results, which may seem obvious for the minimax community on the one hand, for the multiple testing community on the other hand. This review is however useful to join the theories from both communities. In Section 3, we investigate the parallel that can be drawn between aggregation-based minimax adaptive tests and some classical multiple testing procedures, leading to the definitions of the wFWSR and the FWSR for multiple testing procedures. We present general results, as well as a careful study of simple multiple testing problems in Gaussian regression frameworks. Perspectives are finally given in Section 4, and the main proofs are postponed to Section 5.

2. Preliminaries. *Notation*. Let X be an observed random variable taking values in a measurable space $(\mathbb{X}, \mathcal{X})$, whose unknown distribution P belongs to a set $\mathcal{P}$ of possible probability distributions on $(\mathbb{X}, \mathcal{X})$.

Following Goeman and Solari [15], a hypothesis H is a subset of $\mathcal{P}$ and H is true under P if P belongs to H, and false under P otherwise.

Given a finite collection $\mathcal{H}$ of such hypotheses, the aim is simultaneously testing H against $\mathcal{P} \setminus H$, for every H in $\mathcal{H}$, which is equivalent to simultaneously testing "H is true under P" against "H is false under P", or "$P \in H$" against "$P \notin H$", for every H in $\mathcal{H}$.

The sets of true and false hypotheses under P are respectively defined by

$$\mathcal{T}(P) = \{H \in \mathcal{H}, P \in H\} \quad \text{and} \quad \mathcal{F}(P) = \mathcal{H} \setminus \mathcal{T}(P) = \{H \in \mathcal{H}, P \notin H\}.$$

A multiple testing procedure or a multiple test is a statistic given by a collection of rejected hypotheses $\mathcal{R} \subset \mathcal{H}$, only depending on the observed random variable X, whose goal is to infer the set $\mathcal{F}(P)$ of false hypotheses under P in $\mathcal{H}$. In the following, $\cap\mathcal{H}$ is an abbreviation for $\bigcap_{H \in \mathcal{H}} H$.

Most of the tests presented here are based on test statistics, for which the following formalism is needed. For any real valued statistic T, it is classical to consider its cumulative distribution function (c.d.f.) F, which is càdlàg, and its generalized inverse c.d.f. or quantile function F^{-1}, which is càglàd. In the sequel, we also focus on the càglàd c.d.f. of T, F_-, defined by

$$\forall t \in \mathbb{R} \qquad F_-(t) = P(T < t).$$

Its generalized inverse function F_-^{-1} is then a càdlàg function defined by

$$\forall u \in (0, 1) \qquad F_-^{-1}(u) = \sup\{t, F_-(t) \leq u\}.$$

2.1. *Tests of a single null hypothesis and p-values.* A test of a single null hypothesis H_0 is usually formalized as a statistic taking values in $\{0, 1\}$, whose value 1 amounts to rejecting H_0.

There are two classical ways of defining such a test, either by giving a test statistic and the corresponding critical values, or by giving a p-value. The following preliminary result allows us to precisely go back and forth between the "multiple tests" literature used to p-values, and the "aggregated tests" literature, exclusively using test statistics and critical values. Notice that a part of the statements of this result can be proved with Lemma 1.1. in [24]. We nevertheless give a comprehensive and self-contained proof in the supplementary material [13].

LEMMA 1. *Let T be a real-valued test statistic of a single null hypothesis H_0 whose distribution does not depend on P provided that P belongs to H_0. Denote by F and F_- the* (*càdlàg*) *c.d.f. and the càglàd c.d.f. of this distribution under H_0, and by F^{-1} and F_-^{-1} their respective generalized inverse functions as defined above. Let $p(T) = 1 - F_-(T)$, and for any α in* $(0, 1)$,

$$\phi = \mathbb{1}_{\{T > F^{-1}(1-\alpha)\}}, \qquad \phi_- = \mathbb{1}_{\{T > F_-^{-1}(1-\alpha)\}} \quad \textit{and} \quad \phi_p = \mathbb{1}_{\{p(T) \leq \alpha\}}.$$

Then all those three tests are of level α and their associated p-value (*i.e. the limit level α at which they pass from acceptance to rejection*) *is $p(T)$, which satisfies for all P in H_0, $P(p(T) \leq \alpha) \leq \alpha$. Moreover,*

$$T > F_-^{-1}(1-\alpha) \Leftrightarrow p(T) < \alpha \quad \textit{and} \quad \phi_- \leq \phi \wedge \phi_p. \tag{1}$$

Most of the time, c.d.f. are continuous and in this case the three tests ϕ, ϕ_- and ϕ_p are almost surely equal. However, when atoms are present in the distribution, the most powerful one is the test ϕ_p based on the p-value, which is not completely equivalent to the more classical test ϕ based on the test statistic. This can be especially useful when bootstrap or permutation procedures are used since, in this case, Lemma 1 can be applied to the conditional bootstrapped or permuted c.d.f. given the observed random variable, which is naturally noncontinuous (see [29–31] for instance).

Authors using p-values generally consider the test ϕ_p, while authors used to test statistics and critical values generally consider the test ϕ. In order to more conveniently go back and forth between p-values on the one hand, test statistics and critical values on the other hand, regarding the first equivalence stated in (1), we focus all along the paper on tests in the form of

$$\phi_- = \mathbb{1}_{\{T > F_-^{-1}(1-\alpha)\}} = \mathbb{1}_{\{p(T) < \alpha\}}.$$

In particular, when we refer in the sequel to well-known procedures such as Bonferroni or Holm's ones, we in fact refer to the versions of these procedures written in the form of ϕ_- above.

Note that the test ϕ_p can also be expressed using test statistics and critical values (see Corollary 14 in the supplementary material [13]), but at the price of a much more intricate formula.

2.2. *Multiple tests and the Family-Wise Error Rate.* The weak Family-Wise Error Rate of a multiple test $\mathcal{R}$, denoted by wFWER($\mathcal{R}$), is defined by

$$w\mathrm{FWER}(\mathcal{R}) = \sup_{P, \mathcal{T}(P)=\mathcal{H}} P\big(\mathcal{R} \cap \mathcal{T}(P) \neq \varnothing\big). \tag{2}$$

Controlling the wFWER is generally too weak in applications, as some of the hypotheses in $\mathcal{H}$ may actually be false under P. A control of the probability $P(\mathcal{R} \cap \mathcal{T}(P) \neq \varnothing)$, for any possible P, is therefore more appropriate. This leads to the following definition of the (strong) Family-Wise Error Rate of $\mathcal{R}$, denoted by FWER($\mathcal{R}$):

$$\mathrm{FWER}(\mathcal{R}) = \sup_{P \in \mathcal{P}} P\big(\mathcal{R} \cap \mathcal{T}(P) \neq \varnothing\big). \tag{3}$$

Given a prescribed level α in (0, 1), the main concern then becomes to construct a multiple test $\mathcal{R}$ such that

$$\mathrm{FWER}(\mathcal{R}) \leq \alpha, \tag{4}$$

which obviously also implies that $w\mathrm{FWER}(\mathcal{R}) \leq \alpha$.

A large number of multiple tests satisfying (4) have been constructed, among them the historical procedures of Bonferroni and Holm [16, 33], and the more recent min-p type procedures (see [6] for instance). Many of these procedures can be described through the general sequential rejection scheme proposed by Goeman and Solari [15], which consists in iteratively rejecting hypotheses through an application $\mathcal{N}$ from the set of all subsets of $\mathcal{H}$ to itself, as follows:

$$\begin{cases} 1.\ \text{Start with } \mathcal{R}_0 = \varnothing. \\ 2.\ \text{For any } n \geq 0, \text{ build } \mathcal{R}_{n+1} = \mathcal{R}_n \cup \mathcal{N}(\mathcal{R}_n). \\ 3.\ \text{Define } \mathcal{R} = \lim_{n\to\infty} \mathcal{R}_n. \end{cases} \tag{5}$$

Notice that the sequence $(\mathcal{R}_n)_{n\geq 0}$ is always convergent in the present framework since $\mathcal{H}$ is assumed to be finite. For any prescribed α in (0, 1), Goeman and Solari ([15], Theorem 1) proved that sequential rejective procedures satisfy (4), as soon as the two conditions below are true:

$$\text{(Monotonicity)} \qquad \forall \mathcal{S} \subset \mathcal{S}' \subset \mathcal{H} \qquad \mathcal{N}(\mathcal{S}) \subset \mathcal{S}' \cup \mathcal{N}(\mathcal{S}'),$$

$$\text{(Single-Step)} \qquad \forall P \in \mathcal{P} \qquad P\big(\mathcal{N}(\mathcal{F}(P)) \subset \mathcal{F}(P)\big) \geq 1 - \alpha.$$

Let us focus on generic examples, the min-p procedures, assuming that a set $\mathcal{H}$ of hypotheses and their corresponding p-values p_H, for H in $\mathcal{H}$, are given, such that for all P in H, u in (0, 1), $P(p_H \leq u) \leq u$.

For any subset $\mathcal{G}$ of $\mathcal{H}$, and any α in (0, 1), let $q_{\mathrm{mp},\mathcal{G},\alpha}$ be a nonincreasing function of $\mathcal{G}$ such that

$$\forall P \in \cap\mathcal{G} \qquad P\Big(\min_{H\in\mathcal{G}} p_H < q_{\mathrm{mp},\mathcal{G},\alpha}\Big) \leq \alpha.$$

Then a min-p procedure is defined as a sequential rejective procedure with the application $\mathcal{N}$ equal to

$$\mathcal{N}_{\mathrm{mp}} : \mathcal{S} \mapsto \{H \in \mathcal{H} \setminus \mathcal{S}, p_H < q_{\mathrm{mp},\mathcal{H}\setminus\mathcal{S},\alpha}\}.$$

As it satisfies (Monotonicity) and (Single-Step), by [15], Theorem 1, a min-p procedure has a FWER controlled by α.

It is always possible to use $q_{\mathrm{mp},\mathcal{G},\alpha} = \alpha/\#\mathcal{G}$, where $\#\mathcal{G}$ denotes the cardinal of $\mathcal{G}$. The obtained multiple test is due to Holm [16], so we denote it by $\mathcal{R}_{\mathrm{Holm}}$ and the corresponding application by $\mathcal{N}_{\mathrm{Holm}}$. The first step of this procedure corresponds to the well-known Bonferroni multiple test and is denoted by $\mathcal{R}_{\mathrm{Bonf}} := \mathcal{N}_{\mathrm{Holm}}(\varnothing)$.

A more precise choice can be done as follows. If the distribution of $\min_{H\in\mathcal{G}} p_H$ (with c.d.f. $F_{\mathcal{G}}$ and càglàd c.d.f. $F_{\mathcal{G},-}$) does not depend on P in $\cap\mathcal{G}$ and is known, one can take $q_{\mathrm{mp},\mathcal{G},\alpha} = F_{\mathcal{G},-}^{-1}(\alpha)$. The resulting rejection set is then denoted by $\mathcal{R}_{\mathrm{mp}}$. Note that this multiple testing procedure is less conservative than $\mathcal{R}_{\mathrm{Holm}}$, as it contains $\mathcal{R}_{\mathrm{Holm}}$. If $F_{\mathcal{G}}$ is unknown, the quantiles may be replaced by random quantiles, depending on X, based on permutation or bootstrap approaches [29–31], at the possible price of an asymptotic control of the FWER instead of an exact control.

These procedures may be extended to weighted min-p procedures by defining

$$\mathcal{N}_{\mathrm{wmp}} : \mathcal{S} \mapsto \{H \in \mathcal{H} \setminus \mathcal{S}, p_H < w_H q_{\mathrm{wmp},\mathcal{H}\setminus\mathcal{S},\alpha}\},$$

where $(w_H)_{H\in\mathcal{H}}$ is a family of positive weights satisfying $\sum_{H\in\mathcal{H}} w_H = 1$, and where $q_{\mathrm{wmp},\mathcal{G},\alpha}$ satisfies for any α in $(0,1)$,

$$\forall P \in \cap\mathcal{G} \qquad P\Big(\min_{H\in\mathcal{G}} w_H^{-1} p_H < q_{\mathrm{wmp},\mathcal{G},\alpha}\Big) \le \alpha.$$

When the distribution of $\min_{H\in\mathcal{G}} w_H^{-1} p_H$ (with càglàd c.d.f. $F_{w,\mathcal{G},-}$) does not depend on P in $\cap\mathcal{G}$ and is known, one can take $q_{\mathrm{wmp},\mathcal{G},\alpha} = F_{w,\mathcal{G},-}^{-1}(\alpha)$, which defines multiple tests denoted by $\mathcal{R}_{\mathrm{wmp}}$. Note that such procedures are very close to the balanced procedure of Romano and Wolf [31].

2.3. *Aggregated tests and the First Kind Error Rate.* Considering the problem of testing a single null hypothesis H_0 against the alternative $\mathcal{P} \setminus H_0$, we sketch a general methodology for the construction of aggregated tests, and then focus on a classical example.

The idea of aggregated tests comes from the minimax adaptivity theory. Indeed, the construction of minimax adaptive tests, as defined by Spokoiny [35], often consists in the aggregation of a finite collection of initial minimax (nonadaptive) individual tests. In general, a finite collection of hypotheses $\mathcal{H}$ is chosen such that $H_0 \subset \cap\mathcal{H}$, and so that the final aggregated test achieves some expected minimax adaptivity properties. For each hypothesis H in the collection $\mathcal{H}$, an individual test ϕ_H of the null hypothesis H against the alternative $\mathcal{P} \setminus H$ is constructed, that is a statistic with values in $\{0, 1\}$, whose value 1 amounts to rejecting H. The collection of tests is denoted $\Phi_{\mathcal{H}} = \{\phi_H, H \in \mathcal{H}\}$. Then the corresponding aggregated test $\bar{\Phi}_{\mathcal{H}}$ consists in rejecting H_0 if at least one H in $\mathcal{H}$ is rejected with ϕ_H, that is,

$$\bar{\Phi}_{\mathcal{H}} = \sup_{H \in \mathcal{H}} \phi_H. \tag{6}$$

Notice that in the original works, the ϕ_H's are not presented as individual tests of H against $\mathcal{P} \setminus H$, but as some of numerous tests of the original single null hypothesis H_0 against the alternative $\mathcal{P} \setminus H_0$.

Many frameworks have been studied, among them of course Gaussian regression frameworks (see [2, 7, 23, 35] for instance), density or Poisson processes frameworks (see [8, 9, 11, 20, 32]), or more complex ones corresponding to two-sample type problems (see [5, 10, 12]). We mainly focus here on the most simple Gaussian regression framework considered in [1], to illustrate things as clearly as possible.

A Gaussian regression framework.

The observed random variable is a random vector $X = (X_1, \dots, X_n)'$ whose distribution $P = P_f$ is an n-dimensional Gaussian distribution with mean f, and covariance matrix $\sigma^2 I_n$ ($n \geq 1$). The mean $f = (f_1, \dots, f_n)'$ is unknown, while $\sigma^2 > 0$ is assumed to be known.

We consider the problem of testing the single null hypothesis $H_0 = \{P_0\}$ against the alternative $\mathcal{P} \setminus \{P_0\}$, with $\mathcal{P} = \{P_f, f \in \mathbb{R}^n\}$, that is testing "$f = 0$" against "$f \neq 0$."

From a fixed collection $\mathcal{S}$ of vectorial subspaces S of $\mathbb{R}^n$, a collection of tests ϕ_S of H_0 against $\mathcal{P} \setminus H_0$ is constructed, where ϕ_S equals 1 when the norm $\|\Pi_S X\|$ of the orthogonal projection $\Pi_S X$ of X onto S (w.r.t. the Euclidean distance) takes large values. Considering the individual hypothesis $H_S = \{P_f, \Pi_S f = 0\} = \{P_f, f \in S^\perp\}$, ϕ_S may also be viewed as an individual test of H_S against $\mathcal{P} \setminus H_S$, and can thus be denoted by ϕ_{H_S}. For $\mathcal{H} = \{H_S, S \in \mathcal{S}\}$, the aggregated test of the null hypothesis $H_0 = \{P_0\}$ against $\mathcal{P} \setminus \{P_0\}$, based on the collection of tests $\Phi_{\mathcal{H}} = \{\phi_{H_S}, S \in \mathcal{S}\}$ is then defined as in (6) by $\bar{\Phi}_{\mathcal{H}} = \sup_{H_S \in \mathcal{H}} \phi_{H_S}$.

Notice that if the collection $\mathcal{S}$ is not rich enough, then $H_0 \subsetneq \cap \mathcal{H}$.

More concretely, in the following, two different collections of hypotheses are considered, that can be defined from the canonical basis $\{e_1, \dots, e_n\}$ of $\mathbb{R}^n$.

The first one is given by $\mathcal{H} = \{H_{S_i}, i = 1, \dots, n\}$, where for every i in $\{1, \dots, n\}$, $S_i = \mathrm{Vect}(e_i)$ and

$$H_{S_i} = \{P_f, f_i = 0\} = \{P_f, f \in S_i^\perp\}.$$

The second one is given by $\mathcal{H} = \{H_{\bar{S}_i}, i = 1, \dots, n\}$, where for every i in $\{1, \dots, n\}$, $\bar{S}_i = \mathrm{Vect}(e_1, \dots, e_i)$, so

$$H_{\bar{S}_i} = \{P_f, f_1 = \cdots = f_i = 0\} = \{P_f, \Pi_{\bar{S}_i} f = 0\}.$$

Although such collections of nested hypotheses have already been used in the multiple tests framework (as in the second ANOVA problem with ordered alternatives of [26]), they are more usual in the aggregated tests framework. Considering for instance that each $f_i = \langle \mathbf{f}, \varphi_i \rangle$ is the ith coefficient in the expansion of a signal $\mathbf{f}$ on a basis $(\varphi_i)_i$ like the Fourier basis, the testing issue then amounts to detecting the smallest frequency present in the signal.

Note that both collections in particular satisfy $\cap \mathcal{H} = \{P_0\} = H_0$.

First kind error rate of aggregated tests. The first kind error rate of an aggregated test $\bar{\Phi}_{\mathcal{H}}$ of the single null hypothesis H_0 is defined as usual by

$$\mathrm{ER}(\bar{\Phi}_{\mathcal{H}}, H_0) = \sup_{P \in H_0} P(\bar{\Phi}_{\mathcal{H}} = 1) = \sup_{P \in H_0} P\Big(\sup_{H \in \mathcal{H}} \phi_H = 1\Big).$$

This criterion should be controlled by a prescribed level α in $(0, 1)$. For any hypothesis H of the collection $\mathcal{H}$, the individual test ϕ_H is usually defined from a test statistic T_H, whose distribution does not depend on P provided that P belongs to H_0. Respectively denoting by $F_{H,-}$ and $F_{H,-}^{-1}$ the càglàd c.d.f. and càdlàg quantile function of this distribution under H_0, ϕ_H is then defined as $\mathbb{1}_{\{T_H > F_{H,-}^{-1}(1-u_{H,\alpha})\}}$, where $u_{H,\alpha}$ is chosen so that the aggregated test is actually of level α, that is,

$$\mathrm{ER}(\bar{\Phi}_{\mathcal{H}}, H_0) \leq \alpha.$$

The most obvious choice for $u_{H,\alpha}$ is a Bonferroni-type choice $u_{H,\alpha} = \alpha/N$, where $N = \#\mathcal{H}$ is the number of hypotheses in $\mathcal{H}$. This leads to the Bonferroni-type aggregated test $\bar{\Phi}_{\mathcal{H}}^{\mathrm{Bonf}}$ based on the collection

$$\Phi_{\mathcal{H}}^{\mathrm{Bonf}} = \big\{\phi_H^{\mathrm{Bonf}} = \mathbb{1}_{\{T_H > F_{H,-}^{-1}(1-\alpha/N)\}}, H \in \mathcal{H}\big\}.$$

A weighted Bonferroni-type choice $u_{H,\alpha} = w_H \alpha$ is also proposed in [11] and [12], where $(w_H)_{H \in \mathcal{H}}$ is a family of positive weights such that $\sum_{H \in \mathcal{H}} w_H \leq 1$. This leads to the weighted Bonferroni-type aggregated test $\bar{\Phi}_{\mathcal{H}}^{\mathrm{wBonf}}$ based on

$$\Phi_{\mathcal{H}}^{\mathrm{wBonf}} = \big\{\phi_H^{\mathrm{wBonf}} = \mathbb{1}_{\{T_H > F_{H,-}^{-1}(1-w_H\alpha)\}}, H \in \mathcal{H}\big\}.$$

A less conservative choice and still guaranteeing a level α is proposed by Baraud, Huet and Laurent [2]. It consists in taking $u_{H,\alpha} = w_H u_\alpha$, where

$$u_\alpha = \sup\Big\{u, \sup_{P \in H_0} P\big(\exists H \in \mathcal{H}, T_H > F_{H,-}^{-1}(1 - w_H u)\big) \leq \alpha\Big\}.$$

This leads, when $w_H = 1/N$, to the aggregated test $\bar{\Phi}_{\mathcal{H}}^{\mathrm{BHL}}$ based on

$$\Phi_{\mathcal{H}}^{\mathrm{BHL}} = \big\{\phi_H^{\mathrm{BHL}} = \mathbb{1}_{\{T_H > F_{H,-}^{-1}(1-u_\alpha/N)\}}, H \in \mathcal{H}\big\},$$

or, in the general case, to the aggregated test $\bar{\Phi}_{\mathcal{H}}^{\mathrm{wBHL}}$ based on

$$\Phi_{\mathcal{H}}^{\mathrm{wBHL}} = \big\{\phi_H^{\mathrm{wBHL}} = \mathbb{1}_{\{T_H > F_{H,-}^{-1}(1-w_H u_\alpha)\}}, H \in \mathcal{H}\big\}.$$

3. Main results. In this section, we first study the main correspondences between both theories: multiple tests and aggregated tests. To do so, we always assume that a finite collection of hypotheses $\mathcal{H}$ and a single null hypothesis H_0 such that $H_0 \subset \cap\mathcal{H}$ are given.

From any collection $\Phi_{\mathcal{H}} = \{\phi_H, H \in \mathcal{H}\}$ of tests ϕ_H of the single hypothesis H defining an aggregated test, a multiple test of $\mathcal{H}$ is constructed as

$$\mathcal{R}(\Phi_{\mathcal{H}}) = \{H \in \mathcal{H}, \phi_H = 1\}.$$

Conversely, from any multiple test $\mathcal{R}$ of $\mathcal{H}$, we construct

$$\bar{\Phi}(\mathcal{R}) = \mathbb{1}_{\{\mathcal{R} \neq \varnothing\}},$$

which can be seen as an aggregated test of the single null hypothesis H_0.

3.1. *First kind error and first identifications.* First notice that the weak Family-Wise Error Rate of $\mathcal{R}(\Phi_{\mathcal{H}})$ is equal to

$$w\mathrm{FWER}\big(\mathcal{R}(\Phi_{\mathcal{H}})\big) = \sup_{P \in \cap\mathcal{H}} P\big(\mathcal{R}(\Phi_{\mathcal{H}}) \neq \varnothing\big) = \mathrm{ER}(\bar{\Phi}_{\mathcal{H}}, \cap\mathcal{H}).$$

In the same way, one has

$$w\mathrm{FWER}(\mathcal{R}) = \mathrm{ER}\big(\bar{\Phi}(\mathcal{R}), \cap\mathcal{H}\big). \tag{7}$$

Since $H_0 \subset \cap\mathcal{H}$, $w\mathrm{FWER}(\mathcal{R}(\Phi_{\mathcal{H}})) \geq \mathrm{ER}(\bar{\Phi}_{\mathcal{H}}, H_0)$ and $w\mathrm{FWER}(\mathcal{R}) \geq \mathrm{ER}(\bar{\Phi}(\mathcal{R}), H_0)$. Except when $H_0 = \cap\mathcal{H}$, controlling $w\mathrm{FWER}(\mathcal{R}(\Phi_{\mathcal{H}}))$ or $w\mathrm{FWER}(\mathcal{R})$ is thus more difficult than controlling $\mathrm{ER}(\bar{\Phi}_{\mathcal{H}}, H_0)$ or $\mathrm{ER}(\bar{\Phi}(\mathcal{R}), H_0)$, respectively.

Next, assume that for every H in $\mathcal{H}$, a test statistic T_H, whose distribution does not depend on P provided that P belongs to H, is given and denote by p_H its corresponding p-value, as defined by Lemma 1.

PROPOSITION 2. *With the notation of Sections 2.2 and 2.3, the following identifications hold*:

$$\mathcal{R}\big(\Phi_{\mathcal{H}}^{\mathrm{Bonf}}\big) = \mathcal{R}_{\mathrm{Bonf}} \quad \textit{and} \quad \bar{\Phi}_{\mathcal{H}}^{\mathrm{Bonf}} = \bar{\Phi}(\mathcal{R}_{\mathrm{Bonf}}) = \bar{\Phi}(\mathcal{R}_{\mathrm{Holm}}).$$

If additionally the distribution of $\min_{H \in \mathcal{H}} w_H^{-1} p_H$ *does not depend on* P *provided that* P *belongs to* $\cap\mathcal{H}$, *then*

$$\mathcal{N}_{\mathrm{wmp}}(\varnothing) = \mathcal{R}\big(\Phi_{\mathcal{H}}^{\mathrm{wBHL}}\big) \quad \textit{and} \quad \bar{\Phi}_{\mathcal{H}}^{\mathrm{wBHL}} = \bar{\Phi}(\mathcal{R}_{\mathrm{wmp}}).$$

In particular, the first step of the classical min-p procedure is equivalent to the practical procedure introduced by Baraud, Huet and Laurent [2] in the aggregated tests framework.

3.2. *From Separation Rates to Family-Wise Separation Rates*. Let d be a distance on $\mathcal{P}$. For any P in $\mathcal{P}$, and $\mathcal{Q} \subset \mathcal{P}$, we set $d(P, \mathcal{Q}) := \inf_{Q \in \mathcal{Q}} d(P, Q)$.

Separation rates for aggregated tests. Separation rates are second kind error-related quality criteria of a test of $H_0 \subset \mathcal{P}$ against $\mathcal{P} \setminus H_0$. Because $\mathcal{P}$ is in general too large to define separation rates over the whole set $\mathcal{P}$ properly, particularly in nonparametric frameworks, these quantities are first defined on a subset $\mathcal{Q}$ of $\mathcal{P}$. The question of adaptivity with respect to $\mathcal{Q}$ can then be treated. More precisely, we use the following definition due to Baraud [1], which can be viewed as a nonasymptotic version of Ingster's work [19].

DEFINITION 1. Given β in (0, 1), a class of probability distributions $\mathcal{Q} \subset \mathcal{P}$, and a test $\bar{\Phi}$ of a null hypothesis $H_0 \subset \mathcal{P}$, the uniform separation rate of $\bar{\Phi}$ over $\mathcal{Q}$ with prescribed second kind error rate β is defined by

$$\begin{aligned}\mathrm{SR}_d^{\beta}(\bar{\Phi}, \mathcal{Q}, H_0) &= \inf\Big\{r > 0, \sup_{P \in \mathcal{Q}, d(P, H_0) \geq r} P(\bar{\Phi} = 0) \leq \beta\Big\}\\ &= \inf\Big\{r > 0, \inf_{P \in \mathcal{Q}, d(P, H_0) \geq r} P(\bar{\Phi} = 1) \geq 1 - \beta\Big\}.\end{aligned}$$

Note that this definition holds for any null hypothesis, that is any subset of $\mathcal{P}$, and in particular for $\cap\mathcal{H}$. Hence, when $H_0 \subset \cap\mathcal{H}$,

$$\mathrm{SR}_d^{\beta}(\bar{\Phi}, \mathcal{Q}, H_0) \geq \mathrm{SR}_d^{\beta}(\bar{\Phi}, \mathcal{Q}, \cap\mathcal{H}).$$

The corresponding minimax separation rate over $\mathcal{Q}$ with prescribed level α and second kind error rate β is defined as

$$m\mathrm{SR}_d^{\alpha,\beta}(\mathcal{Q}, H_0) = \inf_{\bar{\Phi}, \mathrm{ER}(\bar{\Phi}, H_0) \leq \alpha} \mathrm{SR}_d^{\beta}(\bar{\Phi}, \mathcal{Q}, H_0),$$

where the infimum is taken over all possible level α tests.

A level α test $\bar{\Phi}$ is said to be minimax over $\mathcal{Q}$ if $\mathrm{SR}_d^{\beta}(\bar{\Phi}, \mathcal{Q}, H_0)$ achieves $m\mathrm{SR}_d^{\alpha,\beta}(\mathcal{Q}, H_0)$, possibly up to a multiplicative constant depending on α and β. It is said to be adaptive in the minimax sense over a collection $\mathcal{C}$ of classes $\mathcal{Q}$ if $\mathrm{SR}_d^{\beta}(\bar{\Phi}, \mathcal{Q}, H_0)$ achieves, or nearly achieves, $m\mathrm{SR}_d^{\alpha,\beta}(\mathcal{Q}, H_0)$, for all the classes $\mathcal{Q}$ in $\mathcal{C}$ simultaneously, without knowing in advance the class to which the distribution P belongs.

Weak Family-Wise Separation Rates for multiple tests. Let us now consider a multiple testing procedure $\mathcal{R}$ and the corresponding aggregated test $\bar{\Phi}(\mathcal{R})$. Given β in (0, 1) and a class $\mathcal{Q} \subset \mathcal{P}$, according to Definition 1,

$$\mathrm{SR}_d^{\beta}(\bar{\Phi}(\mathcal{R}), \mathcal{Q}, \cap\mathcal{H}) = \inf\Big\{r > 0, \inf_{P \in \mathcal{Q}, d(P, \cap\mathcal{H}) \geq r} P(\mathcal{R} \neq \varnothing) \geq 1 - \beta\Big\}.$$

This notion is closely related to the maximin optimality criterion considered by Romano, Shaikh and Wolf ([27], Theorem 4.1), whose aim is to find procedures maximizing the power $\inf_{P\in\mathcal{Q}\subset\mathcal{P}\setminus\cap\mathcal{H}} P(\mathcal{R}\neq\varnothing)$. Here, the question is reversed: the aim is to determine a minimal distance r between P (in $\mathcal{Q}$) and $\cap\mathcal{H}$ which guarantees a minimal level of power $(1-\beta)$ for a given procedure. This notion of minimal distance r is considered as a rate of testing (in the spirit of the rates of estimation), and may enable to compare the performance of two testing procedures.

Following the idea of the definition of the weak Family-Wise Error Rate wFWER of $\mathcal{R}$, which is in fact equal to the first kind error rate of $\bar{\Phi}(\mathcal{R})$ for the null hypothesis $\cap\mathcal{H}$ [see (7)], a natural idea would be to define a notion of weak Family-Wise Separation Rate as $\mathrm{SR}_d^{\beta}(\bar{\Phi}(\mathcal{R}),\mathcal{Q},\cap\mathcal{H})$. However, in this second kind error criterion, only alternatives deviating from the intersection $\cap\mathcal{H}$ with a certain distance are taken into account. Considering such a definition would thus amount to confuse multiple tests with their corresponding aggregated tests, seeing all the tested hypotheses as only intermediate hypotheses to an ultimate one: $\cap\mathcal{H}$. This would depart from the multiple testing philosophy, where each tested hypothesis has its own significance and has to be taken into account by itself. In order to address this requirement, instead of alternatives P in $\mathcal{Q}$ such that "$d(P,\cap\mathcal{H})\geq r$", are considered alternatives P in $\mathcal{Q}$ such that "$\exists H\in\mathcal{H}, d(P,H)\geq r$." So, we consider the set $\mathcal{F}_r(P)$ of false hypotheses under P at least at distance r from P, which can be visualized on Figure 1, and which is defined as

$$\mathcal{F}_r(P)=\{H\in\mathcal{H}, d(P,H)\geq r\}.$$

Note in particular that $\mathcal{F}_r(P)\neq\varnothing$ implies that $d(P,\cap\mathcal{H})\geq r$.

DEFINITION 2. Given β in $(0,1)$ and a class of probability distributions $\mathcal{Q}\subset\mathcal{P}$, the weak Family-Wise Separation Rate of a multiple test $\mathcal{R}$ over $\mathcal{Q}$ with prescribed second kind error rate β is defined by

$$\begin{aligned} w\mathrm{FWSR}_d^{\beta}(\mathcal{R},\mathcal{Q}) &= \inf\Big\{r>0, \sup_{P\in\mathcal{Q},\mathcal{F}_r(P)\neq\varnothing} P(\mathcal{R}=\varnothing)\leq\beta\Big\} \\ &= \inf\Big\{r>0, \inf_{P\in\mathcal{Q},\mathcal{F}_r(P)\neq\varnothing} P(\mathcal{R}\neq\varnothing)\geq 1-\beta\Big\}. \end{aligned}$$

The quantity $\pi_1=\inf_{P\in\mathcal{Q},\mathcal{F}_r(P)\neq\varnothing} P(\mathcal{R}\neq\varnothing)$ involved in the above definition is related to $\beta_{\#\mathcal{H},1}(\alpha,r)$ considered in [25], and which can be written here as $\pi_2=\inf_{P\in\mathcal{Q},\mathcal{F}_r(P)\neq\varnothing} P(\mathcal{R}\cap\mathcal{F}(P)\neq\varnothing)$. In [25], Lemmas 3.1 and 4.1, the authors prove that for step-up and step-down procedures, under some assumptions on P, π_1 is equal to π_2, and is maximized by the same step-up and step-down procedures, whatever the distance r. Here, following the spirit of the minimax theory for single tests, we propose to fix a minimal level $1-\beta$ for π_1, and to find which distance r can guarantee its achievement. The notion of weak Family-Wise Separation Rate, thus defined, is linked to the classical one of uniform separation rate thanks to the following result.

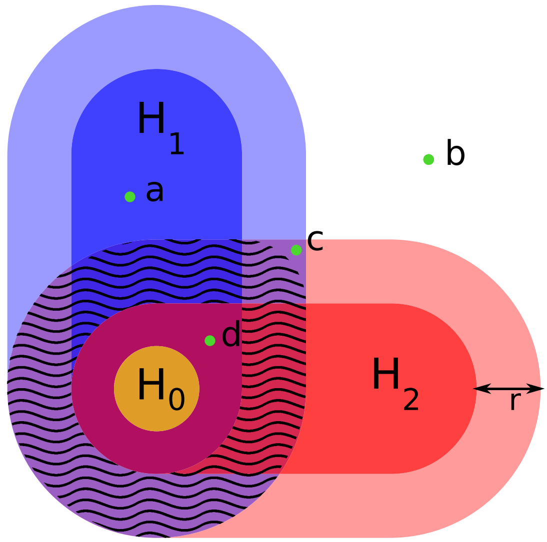

FIG. 1. *Visualization of a multiple testing problem with two hypotheses H_1 and H_2, which are represented with darker colors. Their r-neighborhoods are of lighter shade. The r-neighborhood of $H_1 \cap H_2$ is hatched. A hypothesis H_0 is strictly included in $H_1 \cap H_2$. Point a corresponds to a distribution P such that $\mathcal{T}(P) = \{H_1\}$ and $\mathcal{F}(P) = \mathcal{F}_r(P) = \{H_2\}$. Point b corresponds to a distribution P such that $\mathcal{T}(P) = \varnothing$ and $\mathcal{F}(P) = \mathcal{F}_r(P) = \{H_1, H_2\}$. Point c corresponds to a distribution P such that $\mathcal{T}(P) = \varnothing$, $\mathcal{F}(P) = \{H_1, H_2\}$, $\mathcal{F}_r(P) = \varnothing$ but $d(P, H_1 \cap H_2) \geq r$. Point d corresponds to a distribution P such that $\mathcal{T}(P) = \{H_1, H_2\}$ and $\mathcal{F}(P) = \mathcal{F}_r(P) = \varnothing$ but $P \notin H_0$.*

PROPOSITION 3. *For any subset $\mathcal{Q}$ of $\mathcal{P}$ and β in* $(0, 1)$,

$$w\mathrm{FWSR}_d^\beta(\mathcal{R}, \mathcal{Q}) \leq \mathrm{SR}_d^\beta(\bar{\Phi}(\mathcal{R}), \mathcal{Q}, \cap\mathcal{H}),$$

with an equality if the collection of hypotheses $\mathcal{H}$ and the distance d satisfy

$$\forall r > 0 \qquad \mathcal{F}_r(P) \neq \varnothing \quad \text{if and only if} \quad d(P, \cap\mathcal{H}) \geq r. \tag{8}$$

Looking at Figure 1, it is clear that the above inequality may be strict when condition (8) is not satisfied: for example, the point c is considered in $\mathrm{SR}_d^\beta(\bar{\Phi}(\mathcal{R}), \mathcal{Q}, \cap\mathcal{H})$ but not in $w\mathrm{FWSR}_d^\beta(\mathcal{R}, \mathcal{Q})$. It may therefore be strictly more difficult to control $\mathrm{SR}_d^\beta(\bar{\Phi}(\mathcal{R}), \mathcal{Q}, \cap\mathcal{H})$ than $w\mathrm{FWSR}_d^\beta(\mathcal{R}, \mathcal{Q})$, the alternative set being smaller in the last case.

Note that if the collection $\mathcal{H}$ is closed (under intersection), that is,

$$H \in \mathcal{H} \quad \text{and} \quad H' \in \mathcal{H} \Rightarrow H \cap H' \in \mathcal{H},$$

then condition (8) is always satisfied. For instance, in the Gaussian regression framework considered in Section 2.3, the collection $\mathcal{H} = \{H_{\bar{S}_i}, i = 1, \dots, n\}$ is closed and condition (8) is satisfied.

Furthermore, in the same Gaussian regression framework, with $\mathcal{H} = \{H_{S_i}, i = 1, \dots, n\}$ (which is not closed), if d is taken as $d = d_\infty$, with

$$d_\infty(P_f, P_g) = \|f - g\|_\infty = \max_{i=1,\dots,n} |f_i - g_i|, \tag{9}$$

then condition (8) is also satisfied. But this is not true when using any other distance d_s for $s \geq 1$ defined by

$$d_s(P_f, P_g) = \left(\sum_{i=1}^{n} |f_i - g_i|^s\right)^{1/s}. \tag{10}$$

See also Figure 1 drawn with d_2. In this case, a more general result, based on a more general equivalence property, can be used.

PROPOSITION 4. *Let d be a distance on $\mathcal{P}$, and $\mathcal{Q}$ be a subset of $\mathcal{P}$. If there exists some distance d' on $\mathcal{P}$ such that*

$$\forall P \in \mathcal{Q}, \forall r > 0 \qquad \mathcal{F}_r(P) \neq \varnothing \quad \textit{if and only if} \quad d'(P, \cap\mathcal{H}) \geq r, \tag{11}$$

then for every β in $(0, 1)$,

$$w\mathrm{FWSR}_d^\beta(\mathcal{R}, \mathcal{Q}) = \mathrm{SR}_{d'}^\beta(\bar{\Phi}(\mathcal{R}), \mathcal{Q}, \cap\mathcal{H}).$$

In the above Gaussian regression framework with $\mathcal{H} = \{H_{S_i}, i = 1, \dots, n\}$, if the distance d is equal to any distance d_s ($s \geq 1$) defined by (10), then condition (11) is actually satisfied with $d' = d_\infty$. Thus, for every multiple test $\mathcal{R}$ of $\mathcal{H}$, every subset $\mathcal{Q}$ of $\mathcal{P}$, and every s in $[1, \infty]$,

$$w\mathrm{FWSR}_{d_s}^\beta(\mathcal{R}, \mathcal{Q}) = \mathrm{SR}_{d_\infty}^\beta(\bar{\Phi}(\mathcal{R}), \mathcal{Q}, \{P_0\}).$$

Family-Wise Separation Rates for multiple tests. We now introduce a stronger notion of Family-Wise Separation Rate, which defines a new second kind error-related quality criterion for multiple tests. It allows us to develop a minimax approach in the multiple testing set-up, by bringing it closer to the well developed minimax theory for classical tests of a single null hypothesis.

DEFINITION 3. Given β in $(0, 1)$ and a class of probability distributions $\mathcal{Q} \subset \mathcal{P}$, the Family-Wise Separation Rate of a multiple test $\mathcal{R}$ over $\mathcal{Q}$ with prescribed second kind error rate β is defined by

$$\begin{aligned}\mathrm{FWSR}_d^\beta(\mathcal{R}, \mathcal{Q}) &= \inf\Big\{r > 0, \sup_{P \in \mathcal{Q}} P\big(\mathcal{F}_r(P) \cap (\mathcal{H} \setminus \mathcal{R}) \neq \varnothing\big) \leq \beta\Big\} \\ &= \inf\Big\{r > 0, \inf_{P \in \mathcal{Q}} P\big(\mathcal{F}_r(P) \subset \mathcal{R}\big) \geq 1 - \beta\Big\}.\end{aligned}$$

Remark that computing the Family-Wise Separation Rate requires to control the probability that at least one false hypothesis under P at a distance r from P is accepted. As for the weak Family-Wise Separation Rate, this criterion is thus uniformly valid for each of the tested hypotheses, and does not only make sense for their intersection.

By definition, for fixed $\mathcal{Q}$, $\mathrm{FWSR}_d^{\beta}(\mathcal{R}, \mathcal{Q})$ is monotonous in $\mathcal{R}$, that is,

$$\mathcal{R} \subset \mathcal{R}' \text{ a.s. } \Rightarrow \mathrm{FWSR}_d^{\beta}(\mathcal{R}', \mathcal{Q}) \leq \mathrm{FWSR}_d^{\beta}(\mathcal{R}, \mathcal{Q}). \tag{12}$$

In the same way, for fixed $\mathcal{R}$, $\mathrm{FWSR}_d^{\beta}(\mathcal{R}, \mathcal{Q})$ is monotonous in $\mathcal{Q}$, that is,

$$\mathcal{Q} \subset \mathcal{Q}' \Rightarrow \mathrm{FWSR}_d^{\beta}(\mathcal{R}, \mathcal{Q}) \leq \mathrm{FWSR}_d^{\beta}(\mathcal{R}, \mathcal{Q}').$$

The Family-Wise Separation Rate is naturally a stronger quality criterion than the weak Family-Wise Separation Rate, as stated in the following result.

PROPOSITION 5. *For any distance d, any subset $\mathcal{Q}$ of $\mathcal{P}$, any multiple test $\mathcal{R}$, and β in* $(0, 1)$,

$$w\mathrm{FWSR}_d^{\beta}(\mathcal{R}, \mathcal{Q}) \leq \mathrm{FWSR}_d^{\beta}(\mathcal{R}, \mathcal{Q}).$$

Minimax Family-Wise Separation Rates. Let us now introduce the corresponding minimax approach for multiple tests.

DEFINITION 4. Given α and β in $(0, 1)$, a class of probability distributions $\mathcal{Q} \subset \mathcal{P}$, the minimax Family-Wise Separation Rate over $\mathcal{Q}$ with prescribed FWER α and prescribed second kind error rate β is defined by

$$m\mathrm{FWSR}_d^{\alpha,\beta}(\mathcal{Q}) = \inf_{\mathcal{R}, \mathrm{FWER}(\mathcal{R}) \leq \alpha} \mathrm{FWSR}_d^{\beta}(\mathcal{R}, \mathcal{Q}),$$

where the infimum is taken over all possible multiple tests with a FWER controlled by α.

A multiple test $\mathcal{R}$, whose FWER is controlled by α, is said to be minimax over $\mathcal{Q}$ if $\mathrm{FWSR}_d^{\beta}(\mathcal{R}, \mathcal{Q})$ achieves $m\mathrm{FWSR}_d^{\alpha,\beta}(\mathcal{Q})$, possibly up to a multiplicative constant depending on α and β. It is said to be adaptive in the minimax sense over a collection $\mathcal{C}$ of classes $\mathcal{Q}$ if $\mathrm{FWSR}_d^{\beta}(\mathcal{R}, \mathcal{Q})$ achieves, or nearly achieves, $m\mathrm{FWSR}_d^{\alpha,\beta}(\mathcal{Q})$, for all the classes $\mathcal{Q}$ in $\mathcal{C}$ simultaneously, without knowing in advance the class to which the distribution P belongs.

From the monotonicity properties of FWSR_d^{β}, we deduce that

$$\mathcal{Q} \subset \mathcal{Q}' \Rightarrow m\mathrm{FWSR}_d^{\alpha,\beta}(\mathcal{Q}) \leq m\mathrm{FWSR}_d^{\alpha,\beta}(\mathcal{Q}').$$

It is furthermore to be underlined that when $\mathcal{H}$ is reduced to a single hypothesis H_0, for any multiple test $\mathcal{R}$ and any subset $\mathcal{Q}$ of $\mathcal{P}$,

$$\begin{cases} w\mathrm{FWER}(\mathcal{R}) = \mathrm{FWER}(\mathcal{R}) = \mathrm{ER}(\bar{\Phi}(\mathcal{R}), H_0), \\ w\mathrm{FWSR}_d^{\beta}(\mathcal{R}, \mathcal{Q}) = \mathrm{FWSR}_d^{\beta}(\mathcal{R}, \mathcal{Q}) = \mathrm{SR}_d^{\beta}(\bar{\Phi}(\mathcal{R}), \mathcal{Q}, H_0). \end{cases}$$

Conversely, for any single test $\bar{\Phi}$ of H_0,

$$\begin{cases} \mathrm{ER}(\bar{\Phi}, H_0) = w\mathrm{FWER}(\mathcal{R}(\{\bar{\Phi}\})) = \mathrm{FWER}(\mathcal{R}(\{\bar{\Phi}\})), \\ \mathrm{SR}_d^{\beta}(\bar{\Phi}, \mathcal{Q}, H_0) = w\mathrm{FWSR}_d^{\beta}(\mathcal{R}(\{\bar{\Phi}\}), \mathcal{Q}) = \mathrm{FWSR}_d^{\beta}(\mathcal{R}(\{\bar{\Phi}\}), \mathcal{Q}). \end{cases}$$

Then it is easy to prove that when $\mathcal{H}$ is reduced to a single hypothesis H_0, for any subset $\mathcal{Q}$ of $\mathcal{P}$, $m\mathrm{FWSR}_d^{\alpha,\beta}(\mathcal{Q}) = m\mathrm{SR}_d^{\alpha,\beta}(\mathcal{Q}, H_0)$. In this sense, the present minimax approach for multiple tests can be viewed as a generalization of the classical minimax theory for single hypothesis tests.

When $\mathcal{H}$ is not reduced to a single hypothesis H_0, both theories nevertheless still have, under particular conditions, close links that are established below.

THEOREM 6. *Let d be a distance on $\mathcal{P}$, and $\mathcal{Q}$ be a subset of $\mathcal{P}$. If there exists some distance d' on $\mathcal{P}$ satisfying* (11), *then for every α, β in* $(0, 1)$,

$$m\mathrm{FWSR}_d^{\alpha,\beta}(\mathcal{Q}) \geq m\mathrm{SR}_{d'}^{\alpha,\beta}(\mathcal{Q}, \cap\mathcal{H}). \tag{13}$$

The main role of the previous result is to provide straightforward lower bounds for the minimax Family-Wise Separation Rate over some classes $\mathcal{Q}$ by using the abundant literature on classical minimax testing.

As proved by Theorem 9 in the following, note that some of these lower bounds are tight in some particular Gaussian regression frameworks.

Moreover, as a particular case of the previous result, if (8) holds, then for any subset $\mathcal{Q}$ of $\mathcal{P}$ and α, β in $(0, 1)$,

$$m\mathrm{FWSR}_d^{\alpha,\beta}(\mathcal{Q}) \geq m\mathrm{SR}_d^{\alpha,\beta}(\mathcal{Q}, \cap\mathcal{H}). \tag{14}$$

This formalizes the natural idea that testing multiple hypotheses is more difficult than testing a single hypothesis.

If condition (8) is not satisfied, inequality (14) may not hold either, as shown with the example of Theorem 7.

3.3. *Minimax Family-Wise Separation Rates in the Gaussian regression framework.* In this section, we investigate some (minimax) Family-Wise Separation Rates in the classical Gaussian regression framework presented in Section 2.3, and then in a less common, but still Gaussian, regression framework.

3.3.1. *Classical Gaussian regression framework.* Consider the Gaussian regression framework presented in Section 2.3, with the two collections of hypotheses $\mathcal{H}$, respectively equal to $\{H_{S_i}, i = 1, \dots, n\}$ and $\{H_{\bar{S}_i}, i = 1, \dots, n\}$.

Baraud [1] studies the minimax Separation Rates for the null hypothesis $H_0 = \cap\mathcal{H} = \{P_0\}$ with $d = d_2$, over the classes of alternatives $\mathcal{Q} = \mathcal{P}_k$ defined, for any integer $k \le n$, by

$$\mathcal{P}_k = \{P_f, |f|_0 \le k\}, \tag{15}$$

where $|f|_0$ is the number of nonzero coefficients in f. He proves in particular that for α and β in (0, 1) such that $\alpha + \beta \le 0.5$ and $k \ge 1$,

$$m\mathrm{SR}_{d_2}^{\alpha,\beta}(\mathcal{P}_k, H_0) \ge \sigma\left(k \ln\left(1 + \frac{n}{k^2} \vee \sqrt{\frac{n}{k^2}}\right)\right)^{1/2}, \tag{16}$$

and that this lower bound is tight. Baraud, Huet and Laurent [2] then build aggregated tests that are adaptive over a collection of classes $\mathcal{P}_k$, when σ^2 is unknown, and Laurent, Loubes, Marteau [23] further study the case of heteroscedasticity. In a preliminary version [22], they also prove that

$$m\mathrm{SR}_{d_\infty}^{\alpha,\beta}(\mathcal{P}_k, H_0) \ge \sigma\sqrt{\ln(1 + n)}, \tag{17}$$

by remarking that

$$m\mathrm{SR}_{d_\infty}^{\alpha,\beta}(\mathcal{P}_k, H_0) \ge m\mathrm{SR}_{d_\infty}^{\alpha,\beta}(\mathcal{P}_1, H_0) = m\mathrm{SR}_{d_2}^{\alpha,\beta}(\mathcal{P}_1, H_0),$$

and using Baraud's lower bound.

Multiple testing problem of $\mathcal{H} = \{H_{S_i}, i = 1, \dots, n\}$. For this problem, consider the distance $d = d_s$ (s in $[1, \infty]$), as defined in (9) and (10). Let k be a fixed integer in $\{1, \dots, n\}$. Using $d' = d_\infty$ in Theorem 6 leads to

$$m\mathrm{FWSR}_{d_s}^{\alpha,\beta}(\mathcal{P}_k) \ge m\mathrm{SR}_{d_\infty}^{\alpha,\beta}(\mathcal{P}_k, H_0).$$

From (17), we then obtain that for any α, β in (0, 1) such that $\alpha + \beta \le 0.5$, for any s in $[1, \infty]$, for any k in $\{1, \dots, n\}$,

$$m\mathrm{FWSR}_{d_s}^{\alpha,\beta}(\mathcal{P}_k) \ge \sigma\sqrt{\ln(1 + n)}. \tag{18}$$

Let us now prove that this lower bound is achieved. To do so, let us consider for any $i = 1, \dots, n$, the p-value p_i corresponding to the test that rejects H_{S_i} when $T_i = |X_i|\sigma^{-1} > F^{-1}(1 - \alpha/2)$, where F is here the c.d.f. of a standard Gaussian distribution. Notice that since the Gaussian distribution is continuous, the three tests defined in Lemma 1 are identical, that is,

$$\mathbb{1}_{\{T_i > F^{-1}(1-\alpha/2)\}} = \mathbb{1}_{\{T_i > F_-^{-1}(1-\alpha/2)\}} = \mathbb{1}_{\{p_i \le \alpha\}}.$$

THEOREM 7. *Let α in $(0,1)$, and $\mathcal{R}$ be one of the four multiple testing procedures $\mathcal{R}_{\mathrm{Bonf}}, \mathcal{R}_{\mathrm{Holm}}, \mathcal{R}_{\mathrm{mp}}$ and $\mathcal{R}(\Phi_{\mathcal{H}}^{\mathrm{BHL}})$, based on the p-values p_i defined above, such that* $\mathrm{FWER}(\mathcal{R}) \leq \alpha$. *Then for all s in $[1,\infty]$, k in $\{1,\dots,n\}$, and β in* $(0, 0.5)$,

$$\mathrm{FWSR}_{d_s}^{\beta}(\mathcal{R}, \mathcal{P}_k) \leq \sigma\big(\sqrt{2\ln(k/(2\beta))} + \sqrt{2\ln(n/\alpha)}\big).$$

Comments.

(i) This proves that the four considered multiple testing procedures are minimax over the classes $\mathcal{P}_k$ with a Family-Wise Separation Rate of order $\sigma(\ln n)^{1/2}$, up to a multiplicative constant when $\alpha+\beta \leq 0.5$. Since the considered multiple tests do not depend on the value of k, they are moreover adaptive in the minimax sense over the collection of all the classes $\mathcal{P}_k$, for $k = 1,\dots,n$. Notice that asymptotically, there is here no additional price to pay for adaptation, and that such a phenomenon is rather rarely observed in minimax adaptive testing problems: to our knowledge, only three cases are identified in [11, 12] and [23].

(ii) This also proves at the same time that the minimax Family-Wise Separation Rate over $\mathcal{P}_k$ is of order $\sigma(\ln n)^{1/2}$. By comparison, the minimax Separation Rate $m\mathrm{SR}_{d_2}^{\alpha,\beta}(\mathcal{P}_k, H_0)$ is of order $\sigma n^{\gamma/2}(\ln n)^{1/2}$ when k is proportional to n^{γ} for γ in $(0, 1/2)$ [see (16)], which is much larger than this minimax Family-Wise Separation Rate. This could let think that, when considering the distance $d = d_2$, performing a multiple testing procedure may be much easier than performing a test of a single hypothesis, which would be completely counterintuitive. However, making such a comparison consists of comparing quantities that are not comparable at all. When $d = d_2$, we indeed noticed (see comments below Proposition 4) that the set of alternatives considered in the definition of $w\mathrm{FWSR}_{d_2}^{\beta}(\mathcal{R}, \mathcal{P}_k)$ of any multiple test $\mathcal{R}$ is smaller than the set of alternatives in the definition of $\mathrm{SR}_{d_2}^{\beta}(\bar{\Phi}(\mathcal{R}), \mathcal{P}_k, H_0)$, but exactly equal to the one in $\mathrm{SR}_{d_\infty}^{\beta}(\bar{\Phi}(\mathcal{R}), \mathcal{P}_k, H_0)$. This explains why $m\mathrm{FWSR}_{d_2}^{\alpha,\beta}(\mathcal{P}_k)$, and more generally $m\mathrm{FWSR}_{d_s}^{\alpha,\beta}(\mathcal{P}_k)$ ($s \in [1,\infty]$), are of the same order as the minimax Separation Rate $m\mathrm{SR}_{d_\infty}^{\alpha,\beta}(\mathcal{P}_k, H_0)$ derived in [22].

(iii) Under H_0, the observations are n i.i.d. centered Gaussian variables with variance σ and their maximum is in expectation of order $\sigma(\ln n)^{1/2}$. It is therefore quite expected that no signal with infinite norm smaller than this bound can be detected.

(iv) When $\mathcal{H}$ is reduced to a single hypothesis H_{S_i}, then $m\mathrm{FWSR}_{d_s}^{\alpha,\beta}(\mathcal{P}_k) = m\mathrm{SR}_{d_s}^{\alpha,\beta}(\mathcal{P}_k, H_{S_i})$, both being of order σ. In this sense, $(\ln n)^{1/2}$ can be viewed as the price to pay for multiplicity.

(v) It may be surprising or disappointing that the considered procedures are all minimax adaptive, though we may expect that the Bonferroni one appears as less performing than the three other procedures, constructed through step-down methods, known to give better performance from a second kind error point of view (see [25]). We guess that the gain in Family-Wise Separation Rates of such step-down procedures, probably minor when the Gaussian vector X has such present independent components, is in fact hidden in multiplicative constants. Such a phenomenon is also well known when studying the FWER of Bonferroni, Holm and more general min-p procedures. This gain actually become clearly visible when X is assumed to have a particular dependence structure, as in the example treated in Section 3.3.2.

(vi) If $\alpha \le n/5$ and $\beta \le 0.1$, using [18], Theorem 2.1, instead of (23) in the proof of Theorem 7 would lead to a much sharper bound for moderate size n. However, this argument would deteriorate the bound from an asymptotic point of view.

The study of the present Gaussian framework highlights another interesting point. Baraud's [1] result gives that when $\sqrt{n} \le k \le n$, $m\mathrm{SR}_{d_2}^{\alpha,\beta}(\mathcal{P}_k, H_0)$ is of order $\sigma n^{1/4}$. We prove in the next proposition that although $\mathcal{R}_{\mathrm{Bonf}}$ achieves an optimal $\mathrm{FWSR}_{d_2}^{\beta}$ over any class $\mathcal{P}_k$, its corresponding aggregated test $\bar{\Phi}(\mathcal{R}_{\mathrm{Bonf}})$ does not necessarily satisfy similar minimax optimality properties.

PROPOSITION 8. *For any* α, β *in* $(0, 1)$ *such that* $\alpha \le n/5$ *and* $\alpha + \beta < 1$, *and any* k *such that* $\sqrt{n} \le k \le n$,

$$\begin{aligned}
&\mathrm{SR}_{d_2}^{\beta}(\bar{\Phi}(\mathcal{R}_{\mathrm{Bonf}}), \mathcal{P}_k, H_0) \\
&\quad \ge \sigma\sqrt{k}\left(\sqrt{2\ln(2n/\alpha)} - \frac{\ln(4\ln(2n/\alpha)) + 2}{2\sqrt{2\ln(2n/\alpha)}} - \sqrt{2\ln\bigl(k/(1-\alpha-\beta)\bigr)}\right).
\end{aligned}$$

Notice that the above lower bound for $\mathrm{SR}_{d_2}^{\beta}(\bar{\Phi}(\mathcal{R}_{\mathrm{Bonf}}), \mathcal{P}_k, H_0)$ is of order $\sigma n^{\gamma/2}(\ln n)^{1/2}$ when $k = n^{\gamma}$ with $1/2 \le \gamma < 1$ and n is large enough. Hence, for such a choice of k, with n large, whereas $\mathcal{R}_{\mathrm{Bonf}}$ is minimax over $\mathcal{P}_k$, $\bar{\Phi}(\mathcal{R}_{\mathrm{Bonf}})$ is suboptimal from the minimax point of view over the same $\mathcal{P}_k$.

Conversely, it is easy to show, using the control of non-central chi-square quantiles in [4] for instance, that the test rejecting H_0 when $\sum_{i=1}^n X_i^2 > \sigma^2 q_{(n)}^{1-\alpha}$, where $q_{(n)}^{1-\alpha}$ is the $(1-\alpha)$ quantile of the $\chi^2(n)$ distribution, is minimax over any class $\mathcal{P}_k$. This single test can always be viewed as an aggregated test of n times itself, so that the corresponding multiple test contains all the H_{S_i}'s if $\sum_{i=1}^n X_i^2 > \sigma^2 q_{(n)}^{1-\alpha}$, none of them otherwise. Such a multiple test does not even control its FWER and, therefore, cannot be considered as optimal from the present minimax point of view.

Multiple testing problem of $\mathcal{H} = \{H_{\bar{S}_i}, i = 1, \dots, n\}$. Since the collection $\mathcal{H}$ is closed, condition (8) [or (11) with $d = d'$] is always satisfied, and from Theorem 6, we deduce that for any distance d,

$$m\mathrm{FWSR}_d^{\alpha,\beta}(\mathcal{P}_k) \ge m\mathrm{SR}_d^{\alpha,\beta}(\mathcal{P}_k, H_0).$$

In particular, for $d = d_2$, from (16), the following lower bound is easily derived: for α and β in (0, 1) such that $\alpha + \beta \leq 0.5$, for k in $\{1, \dots, n\}$,

$$m\mathrm{FWSR}_{d_2}^{\alpha,\beta}(\mathcal{P}_k) \geq \sigma\left(k \ln\left(1 + \frac{n}{k^2} \vee \sqrt{\frac{n}{k^2}}\right)\right)^{1/2}. \tag{19}$$

We now introduce a multiple testing procedure, which does not depend on the knowledge of k and whose Family-Wise Separation Rate over $\mathcal{P}_k$ however achieves this lower bound, up to possible multiplicative constants depending on α and β.

As in the above paragraph, let us consider again for any i in $\{1, \dots, n\}$, the p-value p_i associated, thanks to Lemma 1, with the single test that rejects the null hypothesis $H_{S_i} = \{P_f, f_i = 0\}$ when $T_i = |X_i|\sigma^{-1}$ takes large values. We then introduce the multiple test

$$\bar{\mathcal{R}} = \left\{H_{\bar{S}_i}, \min_{j \leq i} p_j \leq \alpha/n\right\}. \tag{20}$$

As the c.d.f. F of the standard Gaussian distribution is continuous,

$$\bar{\mathcal{R}} = \left\{H_{\bar{S}_i}, \max_{j \leq i} T_j > F^{-1}\big(1 - \alpha/(2n)\big)\right\}.$$

This procedure corresponds to a particular basic case of the variant of the closure method of [26] introduced by Romano and Wolf in [29], Algorithm 1 (idealized step-down method), and [29], Theorem 1, when critical values satisfy a monotonicity assumption. In the notation of Romano and Wolf, here $T_{n,i} = \max_{j \leq i} T_j$ and $d_{n,\{1,\dots,i\}} = F^{-1}(1 - \alpha/(2n))$ for all i in $\{1, \dots, n\}$.

THEOREM 9. *Given α in* (0, 1), *let $\bar{\mathcal{R}}$ be the multiple test defined in* (20). *Then* $\mathrm{FWER}(\bar{\mathcal{R}}) \leq \alpha$, *and for any k in* $\{1, \dots, n\}$, *β in* (0, 0.5),

$$\mathrm{FWSR}_{d_2}^{\beta}(\bar{\mathcal{R}}, \mathcal{P}_k) \leq \sigma\sqrt{k}\big(\sqrt{2\ln(n/\alpha)} + \sqrt{-2\ln(2\beta)}\big).$$

Comments.

(i) For k proportional to n^γ with $\gamma \in [0, 1/2)$, notice that this upper bound coincides with the lower bound obtained in (19), up to constants. Hence, in this case at least, $m\mathrm{FWSR}_{d_2}^{\alpha,\beta}(\mathcal{P}_k)$ is of order $\sigma(n^\gamma \ln n)^{1/2}$, so the multiple test $\bar{\mathcal{R}}$ defined by (20) is adaptive in the minimax sense over the considered classes. Notice that there is again here no price to pay for adaptation.

(ii) This result proves that the considered procedure, derived from the sharp variant closure method introduced in [29], but here with basically constant critical values which are a priori not expected to give optimality from a second kind error point of view, is however adaptive in the minimax sense. This may be a bit disturbing. Once again, we guess that the loss in FWSR of such a basic procedure, as compared for instance with sharper procedures involving really monotonous and more precise critical values than α/n for the $\min_{j\leq i} p_j$'s (like the ones actually taken in [29]), is hidden in multiplicative constants. We also guess that this loss would become more visible if the Gaussian vector X had a strong dependence structure, as in the next section.

3.3.2. *Gaussian regression framework with a strong dependence structure.* In this section, we show that Bonferroni procedures are not always optimal and can be outperformed by optimal min-p procedures in the minimax sense. As the gap in FWER between one-step procedures such as Bonferroni ones, and step-down procedures such as min-p ones, is usually more perceptible when the considered p-values are dependent, we here follow this idea, and introduce a somewhat artificial, but determinative, dependent Gaussian regression framework. The chosen dependence structure is quite extreme, so that lower bounds for mFWSRs can be deduced as in the classical minimax theory for single hypothesis tests.

Let $n \geq 1$ and τ be a partition of $\{1, \dots, n\}$. Let $X = (X_1, \dots, X_n)'$ be a random vector whose distribution $P = P_{f,\tau}$ is an n-dimensional Gaussian distribution with mean $f = (f_1, \dots, f_n)'$, and described as follows. For every t in τ, every i in t, $X_i = f_i + \sigma\varepsilon_t$, $(\varepsilon_t)_{t\in\tau}$ being i.i.d. standard Gaussian random variables.

We endow $\mathcal{P} = \{P_{f,\tau}, f \in \mathbb{R}^n, \tau \text{ partition of } \{1, \dots, n\}\}$ with a distance d such that $d(P_{f,\tau}, P_{f',\tau'}) = \|f - f'\|_\infty + \#\tau\Delta\tau'$, where $\tau\Delta\tau'$ denotes the symmetric difference between the partitions τ and τ'.

We consider the collection of hypotheses $\mathcal{H} = \{H_{S_i}, i = 1, \dots, n\}$, with H_{S_i} here defined as $\{P_{f,\tau}, f_i = 0\}$, and for any $i = 1, \dots, n$, the p-value p_i corresponding to the test that rejects H_{S_i} when $T_i = |X_i|\sigma^{-1}$ takes large values, as defined in Lemma 1. Focusing on the class $\mathcal{P}_{k,T}$ defined for T, k in $\{1, \dots, n\}$ by

$$\mathcal{P}_{k,T} = \{P_{f,\tau}, f \in \mathbb{R}^n, |f|_0 \leq k \text{ and } \#\tau = T\},$$

the three following statements hold.

PROPOSITION 10. *For any α, β in $(0, 1)$ such that $\alpha + \beta \leq 0.5$,*

$$m\mathrm{FWSR}_d^{\alpha,\beta}(\mathcal{P}_{k,T}) \geq \sigma\sqrt{\ln T}.$$

PROPOSITION 11. *Let α, β in $(0, 1)$. Let $\mathcal{R}_{\mathrm{mp}}$ and $\mathcal{R}(\Phi_{\mathcal{H}}^{\mathrm{BHL}})$ respectively be the min-p procedure and the multiple test associated with $\Phi_{\mathcal{H}}^{\mathrm{BHL}}$ based on the p-values p_i's defined above, whose* FWER *is controlled by α. Then*

$$\mathrm{FWSR}_d^{\beta}(\mathcal{R}_{\mathrm{mp}}, \mathcal{P}_{k,T}) \leq \mathrm{FWSR}_d^{\beta}(\mathcal{R}(\Phi_{\mathcal{H}}^{\mathrm{BHL}}), \mathcal{P}_{k,T}) \tag{21}$$

and

$$\mathrm{FWSR}_d^{\beta}(\mathcal{R}(\Phi_{\mathcal{H}}^{\mathrm{BHL}}), \mathcal{P}_{k,T}) \leq \sigma\left(\sqrt{2\ln((k \wedge T)/\beta)} + \sqrt{2\ln(T/\alpha)}\right).$$

Note that a sharper result could be obtained for α and β in $(0, 0.2)$ as above, by using the bound for standard Gaussian quantiles of [18], Theorem 2.1.

PROPOSITION 12. *Let* α, β *be fixed levels in* $(0,1)$ *such that* $\alpha \leq n/5$, *and let* $\mathcal{R}_{\mathrm{Bonf}}$ *be the Bonferroni procedure based on the* p*-values* p_i*'s defined above*, *whose* FWER *is controlled by* α. *Then*

$$\begin{aligned}
&\mathrm{FWSR}_d^{\beta}(\mathcal{R}_{\mathrm{Bonf}}, \mathcal{P}_{k,T}) \\
&\qquad \geq \sigma\left(\sqrt{2\ln(2n/\alpha)} - \frac{\ln(4\ln(2n/\alpha))+2}{2\sqrt{2\ln(2n/\alpha)}} - \sqrt{(-2/k)\ln(1-\beta)}\right).
\end{aligned}$$

In particular, as soon as n is large enough, $\mathrm{FWSR}_d^{\beta}(\mathcal{R}_{\mathrm{Bonf}}, \mathcal{P}_{k,T})$ is lower bounded by $\sigma(\ln n)^{1/2}$, up to a multiplicative constant. Therefore, the min-p procedure or the multiple test associated with $\Phi_{\mathcal{H}}^{\mathrm{BHL}}$, that are both not agnostic with respect to the underlying distribution, are able in the present dependent set-up to be minimax over $\mathcal{P}_{k,T}$ with respect to d, and to outperform the Bonferroni procedure as soon as $\ln T \ll \ln n$.

4. Perspectives. The aim of the present work is to lay some foundations of a minimax theory for multiple testing, and in this sense, it has to be viewed as only a starting point for future studies of multiple tests from the minimax point of view.

Lots of emerging issues remain unsolved, encouraging us to pursue this path.

We have proved that the present theory may legitimate one-step and step-down procedures, such as the Bonferroni, Holm or min-p ones for simple multiple testing problems in a very basic Gaussian regression model, where p-values are clearly independent. Our results, and in particular the lower bounds for the minimax Family-Wise Separation Rates, were obtained using classical tools and results from the existing minimax theory for single hypothesis tests. We then have considered another Gaussian regression model, where p-values are roughly dependent, where the Bonferroni procedure is clearly suboptimal from the minimax point of view, contrary to the min-p procedure which is proved to be adaptive in the minimax sense. The present strong dependence structure enables us to use again known results in the classical minimax theory for single hypothesis tests.

Studying some multiple testing problems in other frameworks, typically involving more reasonable dependence structures, will be challenging, all the more as very few works deal with minimax single testing in dependence models. Considering more complex classes of alternatives than the ones introduced here would also be an interesting matter. All this could probably allow to validate already existing sophisticated multiple tests from the second kind error angle, but would also maybe make necessary the construction of new optimal multiple tests.

Questions and problems known to appear in high dimension, which is inherent to many multiple testing problems, will also have to be investigated within the present minimax theory.

Finally, and this is actually closely related to the above question of high dimension, extending the criteria developed here, which are exclusively dedicated to multiple tests controlling the FWER, to multiple tests controlling the False Discovery Rate would be a major progress. It seems to be definitely more difficult, as no parallel between multiple tests controlling the FDR and aggregated tests can be established as clearly as in the present work.

5. Proofs of the main results.

5.1. *Proof of Proposition* 2. The first part of Proposition 2 is a straightforward consequence of (1). Then remark that

$$P\big(\exists H \in \mathcal{H}, T_H > F_{H,-}^{-1}(1 - w_H u)\big) = P\big(\exists H \in \mathcal{H}, w_H^{-1} p_H < u\big) = F_-(u),$$

where F_- is the càglàd c.d.f. of $\min_{H \in \mathcal{H}} w_H^{-1} p_H$, which does not depend on P in $\cap\mathcal{H}$. So $u_\alpha = F_-^{-1}(\alpha)$, and from (1) again, we derive the second part.

5.2. *Proofs of Propositions* 3 *and* 4. If $\mathcal{F}_r(P) \neq \varnothing$, then there exists H in $\mathcal{H}$ such that $d(P, H) \geq r$, that is, such that for any Q in H, $d(P, Q) \geq r$. In particular, this is true for every Q in $\cap\mathcal{H} \subset H$, so $d(P, \cap\mathcal{H}) \geq r$. Therefore,

$$\begin{aligned}&\Big\{r > 0, \sup_{P \in \mathcal{Q}, d(P, \cap\mathcal{H}) \geq r} P(\mathcal{R} = \varnothing) \leq \beta\Big\}\\ &\quad\subset \Big\{r > 0, \sup_{P \in \mathcal{Q}, \mathcal{F}_r(P) \neq \varnothing} P(\mathcal{R} = \varnothing) \leq \beta\Big\},\end{aligned}$$

which leads to the first inequality. But of course under condition (8), both sets are equal and the inequality becomes an equality. This completes the proof of Proposition 3. Proposition 4 is deduced in the same way, just noticing that under condition (11),

$$\begin{aligned}&\Big\{r > 0, \sup_{P \in \mathcal{Q}, d'(P, \cap\mathcal{H}) \geq r} P(\mathcal{R} = \varnothing) \leq \beta\Big\}\\ &\quad= \Big\{r > 0, \sup_{P \in \mathcal{Q}, \mathcal{F}_r(P) \neq \varnothing} P(\mathcal{R} = \varnothing) \leq \beta\Big\}.\end{aligned}$$

5.3. *Proof of Proposition* 5. The result follows from

$$\begin{aligned}&\Big\{r > 0, \sup_{P \in \mathcal{Q}} P\big(\mathcal{F}_r(P) \cap (\mathcal{H} \setminus \mathcal{R}) \neq \varnothing\big) \leq \beta\Big\}\\ &\quad\subset \Big\{r > 0, \sup_{P \in \mathcal{Q}, \mathcal{F}_r(P) \neq \varnothing} P(\mathcal{R} = \varnothing) \leq \beta\Big\},\end{aligned}$$

which is easily obtained by noticing that for $r > 0$,

$$\begin{aligned}\sup_{P \in \mathcal{Q}, \mathcal{F}_r(P) \neq \varnothing} P(\mathcal{R} = \varnothing) &\leq \sup_{P \in \mathcal{Q}, \mathcal{F}_r(P) \neq \varnothing} P\big(\mathcal{F}_r(P) \cap (\mathcal{H} \setminus \mathcal{R}) \neq \varnothing\big)\\ &\leq \sup_{P \in \mathcal{Q}} P\big(\mathcal{F}_r(P) \cap (\mathcal{H} \setminus \mathcal{R}) \neq \varnothing\big).\end{aligned}$$

5.4. *Proof of Theorem* 6. Since for any multiple testing procedure $\mathcal{R}$, $\mathrm{FWER}(\mathcal{R}) \geq w\mathrm{FWER}(\mathcal{R}) = \mathrm{ER}(\bar{\Phi}(\mathcal{R}), \cap\mathcal{H})$ by (7), one has that

$$m\mathrm{FWSR}_d^{\alpha,\beta}(\mathcal{Q}) \geq \inf_{\mathcal{R},\mathrm{ER}(\bar{\Phi}(\mathcal{R}),\cap\mathcal{H})\leq\alpha} \mathrm{FWSR}_d^{\beta}(\mathcal{R}, \mathcal{Q}).$$

By Proposition 5, this leads to

$$m\mathrm{FWSR}_d^{\alpha,\beta}(\mathcal{Q}) \geq \inf_{\mathcal{R},\mathrm{ER}(\bar{\Phi}(\mathcal{R}),\cap\mathcal{H})\leq\alpha} w\mathrm{FWSR}_d^{\beta}(\mathcal{R}, \mathcal{Q}),$$

which is, from Proposition 4, equivalent, when condition (11) is satisfied, to

$$m\mathrm{FWSR}_d^{\alpha,\beta}(\mathcal{Q}) \geq \inf_{\mathcal{R},\mathrm{ER}(\bar{\Phi}(\mathcal{R}),\cap\mathcal{H})\leq\alpha} \mathrm{SR}_{d'}^{\beta}(\bar{\Phi}(\mathcal{R}), \mathcal{Q}, \cap\mathcal{H}).$$

This allows to conclude, as a $\bar{\Phi}(\mathcal{R})$ is a particular single test of $\cap\mathcal{H}$.

5.5. *Proof of Theorem* 7. By construction, $\mathcal{R}_{\mathrm{Bonf}} \subset \mathcal{R}_{\mathrm{Holm}} \subset \mathcal{R}_{\mathrm{mp}}$. By Proposition 2, one also has: $\mathcal{R}_{\mathrm{Bonf}} = \mathcal{N}_{\mathrm{Holm}}(\varnothing) \subset \mathcal{N}_{\mathrm{mp}}(\varnothing) = \mathcal{R}(\Phi_{\mathcal{H}}^{\mathrm{BHL}})$. It is therefore sufficient to upper bound $\mathrm{FWSR}_{d_s}^{\beta}(\mathcal{R}_{\mathrm{Bonf}}, \mathcal{P}_k)$ by (12). So, the aim here is to find a r_0 such that for any $r \geq r_0$ and for any P_f in $\mathcal{P}_k$,

$$P_f\big(\mathcal{F}_r(P_f) \subset \mathcal{R}_{\mathrm{Bonf}}\big) \geq 1 - \beta.$$

By independence, since $d_s(P_f, H_{S_i}) = \inf_{P_g \in H_{S_i}} d_s(P_f, P_g) = |f_i|$,

$$\begin{aligned}
P_f\big(\mathcal{F}_r(P_f) \subset \mathcal{R}_{\mathrm{Bonf}}\big) &= P_f\big(\forall i \text{ s.t. } d_s(P_f, H_{S_i}) \geq r, H_{S_i} \in \mathcal{R}_{\mathrm{Bonf}}\big) \\
&= P_f\big(\forall i \text{ s.t. } |f_i| \geq r, p_i \leq \alpha/n\big) \\
&= \prod_{i,|f_i|\geq r} P_f(p_i \leq \alpha/n).
\end{aligned}$$

Moreover, denoting by F the c.d.f. of a standard Gaussian variable, and by $(\varepsilon_1, \dots, \varepsilon_n)$ a sample of n i.i.d. standard Gaussian variables such that $X_i = f_i + \sigma\varepsilon_i$, for all i in $\{1, \dots, n\}$, recall that

$$p_i = 2F\big(-\sigma^{-1}|X_i|\big) = 2F\big(-|f_i/\sigma + \varepsilon_i|\big).$$

One can easily prove that for all real numbers a, b,

$$F\big(|a| - b\big) \leq \mathbb{P}\big(|a + \varepsilon_i| > b\big) \leq 2F\big(|a| - b\big). \tag{22}$$

Therefore,

$$\begin{aligned}
P_f\big(\mathcal{F}_r(P_f) \subset \mathcal{R}_{\mathrm{Bonf}}\big) &= \prod_{i,|f_i|\geq r} P_f\left(\left|\frac{f_i}{\sigma} + \varepsilon_i\right| \geq -F^{-1}\left(\frac{\alpha}{2n}\right)\right) \\
&\geq \prod_{i,|f_i|\geq r} \left(F\left(\frac{|f_i|}{\sigma} + F^{-1}\left(\frac{\alpha}{2n}\right)\right)\right) \\
&\geq \left(F\left(\frac{r}{\sigma} + F^{-1}\left(\frac{\alpha}{2n}\right)\right)\right)^{\#\mathcal{F}_r(P_f)}.
\end{aligned}$$

Hence, $P_f(\mathcal{F}_r(P_f) \subset \mathcal{R}_{\mathrm{Bonf}}) \geq 1-\beta$ if

$$r \geq \sigma\bigl(F^{-1}\bigl((1-\beta)^{1/\#\mathcal{F}_r(P_f)}\bigr) - F^{-1}\bigl(\alpha/(2n)\bigr)\bigr).$$

Note that for u in (0, 1) and δ in [0, 1], then $-F^{-1}(u) = F^{-1}(1-u)$, and

$$F^{-1}\bigl((1-u)^{\delta}\bigr) \leq F^{-1}(1-\delta u).$$

Therefore, $P_f(\mathcal{F}_r(P_f) \subset \mathcal{R}_{\mathrm{Bonf}}) \geq 1-\beta$ if

$$r \geq \sigma\bigl(F^{-1}\bigl(1-\beta/\#\mathcal{F}_r(P_f)\bigr) + F^{-1}\bigl(1-\alpha/(2n)\bigr)\bigr).$$

Since for every $u > 0$, $1 - F(u) \leq e^{-u^2/2}/2$,

$$\forall u \in (0, 0.5) \qquad F^{-1}(1-u) \leq \sqrt{-2\ln(2u)}. \tag{23}$$

Finally, $P_f(\mathcal{F}_r(P_f) \subset \mathcal{R}_{\mathrm{Bonf}}) \geq 1-\beta$ (with β in $(0, 0.5)$) if

$$r \geq \sigma\bigl(\sqrt{2\ln\bigl(\#\mathcal{F}_r(P_f)/(2\beta)\bigr)} + \sqrt{2\ln(n/\alpha)}\bigr).$$

This gives the result, as $\#\mathcal{F}_r(P_f) \leq k$ when P_f belongs to $\mathcal{P}_k$.

5.6. *Proof of Proposition* 8. Let f be the vector such that $f_i = r/\sqrt{k}$ for $i = 1, \ldots, k$ and $f_i = 0$ for $i \geq k+1$. Then $d_2(P_f, P_0) = r$. We want to choose r such that

$$P_f(\exists i, H_{S_i} \text{ is rejected}) < 1-\beta.$$

With the notation of the above proof of Theorem 7, a union bound gives

$$\begin{aligned}
P_f(\exists i, H_{S_i} \text{ is rejected}) &\leq \sum_{i=1}^{n} P_f(H_{S_i} \text{ is rejected}) \\
&\leq (n-k)\frac{\alpha}{n} + kP_f\left(\left|\frac{r}{\sigma\sqrt{k}} + \varepsilon_i\right| \geq F^{-1}\left(1-\frac{\alpha}{2n}\right)\right) \\
&\leq \alpha + 2kF\left(\frac{r}{\sigma\sqrt{k}} - F^{-1}\left(1-\frac{\alpha}{2n}\right)\right),
\end{aligned}$$

where the last inequality comes from (22).

Hence, we only have to choose r such that

$$r < \sigma\sqrt{k}\bigl(F^{-1}\bigl(1-\alpha/(2n)\bigr) - F^{-1}\bigl(1-(1-\alpha-\beta)/(2k)\bigr)\bigr).$$

From (23), we get

$$-F^{-1}\bigl(1-(1-\alpha-\beta)/(2k)\bigr) \geq -\sqrt{2\ln\bigl(k/(1-\alpha-\beta)\bigr)}.$$

From the lower bound of [18], Theorem 2.1, we deduce that when $\alpha \leq n/5$,

$$F^{-1}\bigl(1-\alpha/(2n)\bigr) \geq \sqrt{2\ln(2n/\alpha)} - \frac{\ln(4\ln(2n/\alpha))+2}{2\sqrt{2\ln(2n/\alpha)}}.$$

Hence, $P_f(\exists i, H_{S_i} \text{ is rejected}) < 1-\beta$ as soon as

$$r < \sigma\sqrt{k}\left(\sqrt{2\ln(2n/\alpha)} - \frac{\ln(4\ln(2n/\alpha))+2}{2\sqrt{2\ln(2n/\alpha)}} - \sqrt{2\ln\bigl(k/(1-\alpha-\beta)\bigr)}\right).$$

This completes the proof.

5.7. *Proof of Theorem* 9. Let us first prove that $\mathrm{FWER}(\bar{\mathcal{R}}) \le \alpha$. For any f, let i_0 be the largest integer i such that $f_1 = \cdots = f_i = 0$. Then

$$\begin{aligned} P_f\bigl(\bar{\mathcal{R}} \cap \mathcal{T}(P_f) \neq \varnothing\bigr) &= P_f(\exists i \le i_0, \exists j \le i, p_j \le \alpha/n) \\ &\le P_f(\exists j \le i_0, p_j \le \alpha/n) \\ &\le \alpha. \end{aligned}$$

Considering $d = d_2$, the goal is now to find a positive real number r_0 such that for any $r \ge r_0$ and for any P_f in $\mathcal{P}_k$,

$$P_f\bigl(\mathcal{F}_r(P_f) \subset \bar{\mathcal{R}}\bigr) \ge 1-\beta.$$

Assume that P_f belongs to $\mathcal{P}_k$. Given $r > 0$,

$$\mathcal{F}_r(P_f) = \{H_{\bar{S}_i}, d_2(P_f, H_{\bar{S}_i}) \ge r\} = \left\{H_{\bar{S}_i}, \sum_{j\in\{1,\dots,i\}} f_j^2 \ge r^2\right\}.$$

Then, if $\sum_{j=1}^n f_j^2 < r^2$, $P_f(\mathcal{F}_r(P_f) \subset \bar{\mathcal{R}}) = 1$. Otherwise, let i_0 be now the smallest integer in $\{1,\dots,n\}$ such that $\sum_{j=1}^{i_0} f_j^2 \ge r^2$. As this sum has at most $i_0 \wedge k$ nonzero terms, there exists j_0 in $\{1,\dots,i_0\}$ such that $f_{j_0}^2 \ge r^2/(i_0 \wedge k) \ge r^2/k$. Furthermore,

$$P_f\bigl(\mathcal{F}_r(P_f) \subset \bar{\mathcal{R}}\bigr) = P_f\left(\forall i \text{ s.t. } \sum_{j\in\{1,\dots,i\}} f_j^2 \ge r^2, \min_{j\in\{1,\dots,i\}} p_j \le \alpha/n\right).$$

If $p_{j_0} \le \alpha/n$, then $\min_{j=1,\dots,i_0} p_j \le \alpha/n$, and for every i in $\{1,\dots,n\}$ such that $\sum_{j=1}^i f_j^2 \ge r^2$, one has that $i \ge i_0 \ge j_0$, and $\min_{j=1,\dots,i} p_j \le \alpha/n$.

The event $\{\forall i \text{ s.t. } \sum_{j=1}^i f_j^2 \ge r^2, \min_{j=1,\dots,i} p_j \le \alpha/n\}$ thus contains the event $\{p_{j_0} \le \alpha/n\}$. Hence, with the notation of the proof of Theorem 7,

$$\begin{aligned} P_f\bigl(\mathcal{F}_r(P_f) \subset \bar{\mathcal{R}}\bigr) &\ge P_f(p_{j_0} \le \alpha/n) \\ &\ge P_f\bigl(2F\bigl(-\sigma^{-1}|X_{j_0}|\bigr) \le \alpha/n\bigr) \\ &\ge P_f\bigl(2F\bigl(-|f_{j_0}/\sigma + \varepsilon_i|\bigr) \le \alpha/n\bigr). \end{aligned}$$

By (22), it follows that

$$\begin{aligned} P_f\bigl(\mathcal{F}_r(P_f) \subset \bar{\mathcal{R}}\bigr) &\ge F\bigl(|f_{j_0}|/\sigma + F^{-1}\bigl(\alpha/(2n)\bigr)\bigr) \\ &\ge F\bigl(r/(\sqrt{k}\sigma) + F^{-1}\bigl(\alpha/(2n)\bigr)\bigr). \end{aligned}$$

Therefore, $P_f(\mathcal{F}_r(P_f) \subset \bar{\mathcal{R}}) \geq 1-\beta$ as soon as

$$r \geq \sigma\sqrt{k}\big(F^{-1}\big(1-\alpha/(2n)\big) + F^{-1}(1-\beta)\big).$$

Finally, by (23), we derive that $P_f(\mathcal{F}_r(P_f) \subset \bar{\mathcal{R}}) \geq 1-\beta$ ($\beta \in (0, 0.5)$) as soon as

$$r \geq \sigma\sqrt{k}\big(\sqrt{2\ln(n/\alpha)} + \sqrt{-2\ln(2\beta)}\big).$$

This completes the proof.

5.8. *Proof of Proposition* 10. The result is clearly satisfied when $T = 1$.

Let us now consider the case where $T \geq 2$ and $k \leq T-1$. We introduce the partition $\tau = \{\{1\}, \dots, \{T-1\}, \{T, \dots, n\}\}$, with T elements where the $T-1$ first elements are singletons. The class $\mathcal{P}_{k,T}$ then contains

$$\mathcal{P}^*_{k,\tau} = \{P_{f,\tau}, \#\{j \leq T-1, f_j \neq 0\} \leq k \text{ and } f_T = \cdots = f_n = 0\},$$

and as $m\mathrm{FWSR}_d^{\alpha,\beta}$ is increasing,

$$\begin{aligned} &m\mathrm{FWSR}_d^{\alpha,\beta}(\mathcal{P}_{k,T}) \\ &\quad \geq m\mathrm{FWSR}_d^{\alpha,\beta}(\mathcal{P}^*_{k,\tau}) \\ &\quad = \inf_{\mathcal{R}, \mathrm{FWER}(\mathcal{R}) \leq \alpha} \inf\Big\{r > 0, \sup_{P_{f,\tau} \in \mathcal{P}^*_{k,\tau}} P_{f,\tau}\big(\mathcal{F}_r(P_{f,\tau}) \cap (\mathcal{H} \setminus \mathcal{R}) \neq \varnothing\big) \leq \beta\Big\}. \end{aligned}$$

Since, for every $P_{f,\tau}$ in $\mathcal{P}^*_{k,\tau}$ and for every $i \geq T$, $f_i = 0$, with any $\mathcal{R}$ such that $\mathrm{FWER}(\mathcal{R}) \leq \alpha$, we can associate $\bar{\mathcal{R}} = \mathcal{R} \cap \{H_{S_i}, i = 1, \dots, T-1\}$. Naturally $\mathrm{FWER}(\bar{\mathcal{R}})$ is also upper bounded by α and $\{\mathcal{F}_r(P_{f,\tau}) \cap (\mathcal{H} \setminus \mathcal{R}) \neq \varnothing\} = \{\mathcal{F}_r(P_{f,\tau}) \cap (\mathcal{H} \setminus \bar{\mathcal{R}}) \neq \varnothing\}$. Hence,

$$P_{f,\tau}\big(\mathcal{F}_r(P_{f,\tau}) \cap (\mathcal{H} \setminus \mathcal{R}) \neq \varnothing\big) = P_{f,\tau}\big(\mathcal{F}_r(P_{f,\tau}) \cap (\mathcal{H} \setminus \bar{\mathcal{R}}) \neq \varnothing\big).$$

Therefore,

$$\begin{aligned} &\inf_{\mathcal{R}, \mathrm{FWER}(\mathcal{R}) \leq \alpha} \inf\Big\{r > 0, \sup_{P_{f,\tau} \in \mathcal{P}^*_{k,\tau}} P_{f,\tau}\big(\mathcal{F}_r(P_{f,\tau}) \cap (\mathcal{H} \setminus \mathcal{R}) \neq \varnothing\big) \leq \beta\Big\} \\ &\quad = \inf_{\substack{\bar{\mathcal{R}}, \mathrm{FWER}(\bar{\mathcal{R}}) \leq \alpha \\ \bar{\mathcal{R}} \cap \{H_{S_i}, i \geq T\} = \varnothing}} \inf\Big\{r > 0, \sup_{P_{f,\tau} \in \mathcal{P}^*_{k,\tau}} P_{f,\tau}\big(\mathcal{F}_r(P_{f,\tau}) \cap (\mathcal{H} \setminus \bar{\mathcal{R}}) \neq \varnothing\big) \leq \beta\Big\}. \end{aligned}$$

Now introduce the set of distributions

$$\tilde{\mathcal{P}}_{k,T} = \{P_f = \mathcal{N}\big(f, \sigma^2 I_{T-1}\big), f \in \mathbb{R}^{T-1}, |f|_0 \leq k\}.$$

With $P_{f,\tau}$ in $\mathcal{P}^*_{k,\tau}$ we associate $P_{\tilde{f}}$ in $\tilde{\mathcal{P}}_{k,T}$ such that for $j \leq T-1$, $\tilde{f}_j = f_j$.

Moreover, consider the hypotheses $(\tilde{H}_{S_j})_{j=1,\dots,T-1}$, subsets of $\tilde{\mathcal{P}}_{k,T}$, defined by $\tilde{H}_{S_j} = \{P_{\tilde{f}}, \tilde{f}_j = 0\}$. When $P_{f,\tau} \in \mathcal{P}^*_{k,\tau}$, it is clear that, for $j \leq T-1$ $d(P_{f,\tau}, H_{S_j}) = d_\infty(P_{\tilde{f}}, \tilde{H}_{S_j})$.

With any $\bar{\mathcal{R}}$, we can associate $\tilde{\mathcal{R}}$, a multiple test of the collection of hypotheses $\tilde{\mathcal{H}} = \{\tilde{H}_{S_j}, j = 1, \dots, T-1\}$ which contains $\tilde{H}_{S_j}$ when $\bar{\mathcal{R}}$ contains H_{S_j}. Hence,

$$\inf_{\substack{\bar{\mathcal{R}}, \mathrm{FWER}(\bar{\mathcal{R}}) \le \alpha \\ \bar{\mathcal{R}} \cap \{H_{S_i}, i \ge T\} = \varnothing}} \inf\Big\{r > 0, \sup_{P_{f,\tau} \in \mathcal{P}^*_{k,\tau}} P_{f,\tau}\big(\mathcal{F}_r(P_{f,\tau}) \cap (\mathcal{H} \setminus \bar{\mathcal{R}}) \ne \varnothing\big) \le \beta\Big\}$$

$$= \inf_{\substack{\tilde{\mathcal{R}} \text{ associated with a} \\ \bar{\mathcal{R}}, \mathrm{FWER}(\bar{\mathcal{R}}) \le \alpha}} \inf\Big\{r > 0, \sup_{P_{\tilde{f}} \in \tilde{\mathcal{P}}_{k,T}} P_{\tilde{f}}\big(\mathcal{F}_r(P_{\tilde{f}}) \cap (\tilde{\mathcal{H}} \setminus \tilde{\mathcal{R}}) \ne \varnothing\big) \le \beta\Big\}$$

$$\ge \inf_{\mathcal{R}, \mathrm{FWER}(\mathcal{R}) \le \alpha} \inf\Big\{r > 0, \sup_{P_{\tilde{f}} \in \tilde{\mathcal{P}}_{k,T}} P_{\tilde{f}}\big(\mathcal{F}_r(P_{\tilde{f}}) \cap (\tilde{\mathcal{H}} \setminus \mathcal{R}) \ne \varnothing\big) \le \beta\Big\},$$

$\mathcal{F}_r(P_{\tilde{f}})$ being defined w.r.t. d_∞, and the last inequality being a consequence of $\mathrm{FWER}(\tilde{\mathcal{R}}) = \mathrm{FWER}(\bar{\mathcal{R}})$. Therefore, by (18),

$$m\mathrm{FWSR}_d^{\alpha,\beta}(\mathcal{P}_{k,T}) \ge m\mathrm{FWSR}_d^{\alpha,\beta}(\mathcal{P}^*_{k,\tau}) \ge m\mathrm{FWSR}_{d_\infty}^{\alpha,\beta}(\tilde{\mathcal{P}}_{k,T}) \ge \sigma\sqrt{\ln T}.$$

Finally, in the case where $T \ge 2$ and $k \ge T$, $\mathcal{P}_{k,T}$ contains $\mathcal{P}_{T-1,T}$, so

$$m\mathrm{FWSR}_d^{\alpha,\beta}(\mathcal{P}_{k,T}) \ge m\mathrm{FWSR}_d^{\alpha,\beta}(\mathcal{P}_{T-1,T}) \ge \sigma\sqrt{\ln T}.$$

5.9. *Proof of Proposition* 11. Inequality (21) is just a consequence of the fact that $\mathcal{R}(\Phi_{\mathcal{H}}^{\mathrm{BHL}}) \subset \mathcal{R}_{\mathrm{mp}}$. Moreover, $\mathcal{R}(\Phi_{\mathcal{H}}^{\mathrm{BHL}})$ contains every hypothesis H_{S_i} such that the p-value $p_i = 2F(-|X_i|/\sigma)$ is smaller than the α quantile of $\min_{i=1,\dots,n} p_i = \min_{i=1,\dots,n}(2F(-|X_i|/\sigma))$ under the distribution P_0. Since when $f = 0$, $\min_{i=1,\dots,n} p_i = \min_{t \in \tau} 2F(-|\varepsilon_t|)$, $\mathcal{R}(\Phi_{\mathcal{H}}^{\mathrm{BHL}})$ contains $\mathcal{R}_{\mathrm{Bonf}}^{\alpha/T} = \{H_{S_i}, p_i \le \alpha/T\}$.

For any f in $\mathbb{R}^n$ and any partition τ of $\{1, \dots, n\}$, $d(P_{f,\tau}, H_{S_i}) = |f_i|$. Therefore, $H_{S_i} \in \mathcal{F}_r(P_{f,\tau})$ if and only if $|f_i| \ge r$, and

$$\begin{aligned} &P_{f,\tau}\big(\mathcal{F}_r(P_{f,\tau}) \subset \mathcal{R}_{\mathrm{Bonf}}^{\alpha/T}\big) \\ &\quad = \prod_{t \in \tau} P_{f,\tau}\big(\forall i \in t \text{ s.t. } |f_i| \ge r, |f_i/\sigma + \varepsilon_t| \ge -F^{-1}(\alpha/(2T))\big). \end{aligned}$$

Hence,

$$P_{f,\tau}\big(\mathcal{F}_r(P_{f,\tau}) \subset \mathcal{R}_{\mathrm{Bonf}}^{\alpha/T}\big) \ge \prod_{t \in \tau, \exists i \in t, |f_i| \ge r} P_{f,\tau}\big(F^{-1}(\alpha/(2T)) + r/\sigma \ge |\varepsilon_t|\big).$$

Then, if $P_{f,\tau}$ belongs to $\mathcal{P}_{k,T}$,

$$P_{f,\tau}\big(\mathcal{F}_r(P_{f,\tau}) \subset \mathcal{R}_{\mathrm{Bonf}}^{\alpha/T}\big) \ge \left(2F\left(\frac{r}{\sigma} + F^{-1}\left(\frac{\alpha}{2T}\right)\right) - 1\right)^{k \wedge T}.$$

Thus, $P_{f,\tau}(\mathcal{F}_r(P_{f,\tau}) \subset \mathcal{R}_{\mathrm{Bonf}}^{\alpha/T}) \ge 1 - \beta$ if

$$r \ge \sigma\left(F^{-1}\left(\frac{1 + (1-\beta)^{1/(k \wedge T)}}{2}\right) + F^{-1}\left(1 - \frac{\alpha}{2T}\right)\right).$$

Note that for any u in (0, 1), δ in [0, 1],

$$F^{-1}\left(\frac{1+(1-u)^{\delta}}{2}\right) \le F^{-1}\left(1-\frac{\delta u}{2}\right).$$

Therefore, $P_{f,\tau}(\mathcal{F}_r(P_{f,\tau}) \subset \mathcal{R}_{\mathrm{Bonf}}^{\alpha/T}) \ge 1-\beta$ if

$$r \ge \sigma\big(F^{-1}(1-\beta/\big(2(k \wedge T)\big)) + F^{-1}(1-\alpha/(2T))\big).$$

By using (23) again, we obtain that

$$\begin{aligned} \mathrm{FWSR}_d^{\beta}\big(\mathcal{R}(\Phi_{\mathcal{H}}^{\mathrm{BHL}}), \mathcal{P}_{k,T}\big) &\le \mathrm{FWSR}_d^{\beta}\big(\mathcal{R}_{\mathrm{Bonf}}^{\alpha/T}, \mathcal{P}_{k,T}\big) \\ &\le \sigma\big(\sqrt{2\ln(k \wedge T/\beta)} + \sqrt{2\ln(T/\alpha)}\big). \end{aligned}$$

5.10. *Proof of Proposition* 12. Let us take $\tau = \{\{1\}, \dots, \{T-1\}, \{T, \dots, n\}\}$ as in the proof of Proposition 10. Take f such that $f_1 = \dots = f_k = r$ and $f_i = 0$ for $i \ge k+1$. Then

$$\begin{aligned} P_{f,\tau}\big(\mathcal{F}_r(P_{f,\tau}) \subset \mathcal{R}_{\mathrm{Bonf}}\big) &= \prod_{i=1}^{k} P_{f,\tau}(H_{S_i} \in \mathcal{R}_{\mathrm{Bonf}}) \\ &= \prod_{i=1}^{k} P_{f,\tau}\left(\left|\frac{r}{\sigma}+\varepsilon_i\right| \ge F^{-1}\left(1-\frac{\alpha}{2n}\right)\right) \\ &\le 2^k F\left(\frac{r}{\sigma} - F^{-1}\left(1-\frac{\alpha}{2n}\right)\right)^k. \end{aligned}$$

Therefore, $P_{f,\tau}(\mathcal{F}_r(P_{f,\tau}) \subset \mathcal{R}_{\mathrm{Bonf}}) < 1-\beta$, if

$$r < \sigma\big(F^{-1}(1-\alpha/(2n)) + F^{-1}((1-\beta)^{1/k}/2)\big),$$

and we conclude as in the proof of Proposition 8.

Acknowledgments. We are very grateful to both referees and the Associate Editor, whose questions and comments allowed us to improve the present paper. We also warmly thank Nicolas Verzelen for his particularly relevant suggestions and for the fruitful discussions about future works.

SUPPLEMENTARY MATERIAL

Additional results and proofs for "Family-Wise Separation Rates for multiple testing" (DOI: 10.1214/15-AOS1418SUPP; .pdf). This supplement contains additional results on the cumulative distribution functions and the proof of Lemma 1.

REFERENCES

[1] BARAUD, Y. (2002). Non-asymptotic minimax rates of testing in signal detection. *Bernoulli* **8** 577–606. MR1935648

[2] BARAUD, Y., HUET, S. and LAURENT, B. (2003). Adaptive tests of linear hypotheses by model selection. *Ann. Statist.* **31** 225–251. MR1962505

[3] BENJAMINI, Y. and HOCHBERG, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **57** 289–300. MR1325392

[4] BIRGÉ, L. (2001). An alternative point of view on Lepski's method. In *State of the Art in Probability and Statistics* (*Leiden*, 1999). *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **36** 113–133. IMS, Beachwood, OH. MR1836557

[5] COLLIER, O. and DALALYAN, A. S. (2015). Curve registration by nonparametric goodness-of-fit testing. *J. Statist. Plann. Inference* **162** 20–42. MR3323102

[6] DUDOIT, S. and VAN DER LAAN, M. J. (2008). *Multiple Testing Procedures with Applications to Genomics*. Springer, New York. MR2373771

[7] DÜMBGEN, L. and SPOKOINY, V. G. (2001). Multiscale testing of qualitative hypotheses. *Ann. Statist.* **29** 124–152. MR1833961

[8] DÜMBGEN, L. and WALTHER, G. (2008). Multiscale inference about a density. *Ann. Statist.* **36** 1758–1785. MR2435455

[9] FROMONT, M. and LAURENT, B. (2006). Adaptive goodness-of-fit tests in a density model. *Ann. Statist.* **34** 680–720. MR2281881

[10] FROMONT, M., LAURENT, B., LERASLE, M. and REYNAUD-BOURET, P. (2012). Kernels based tests with non-asymptotic bootstrap approaches for two-sample problems. In *Journal of Machine Learning Research*: *Workshop and Conference Proceedings*, *COLT* 2012 **23** 23.1–23.22. Edimbourg, UK.

[11] FROMONT, M., LAURENT, B. and REYNAUD-BOURET, P. (2011). Adaptive tests of homogeneity for a Poisson process. *Ann. Inst. Henri Poincaré Probab. Stat.* **47** 176–213. MR2779402

[12] FROMONT, M., LAURENT, B. and REYNAUD-BOURET, P. (2013). The two-sample problem for Poisson processes: Adaptive tests with a nonasymptotic wild bootstrap approach. *Ann. Statist.* **41** 1431–1461. MR3113817

[13] FROMONT, M., LERASLE, M. and REYNAUD-BOURET, P.(2016). Supplement to "Family-Wise Separation Rates for multiple testing." DOI:10.1214/15-AOS1418SUPP.

[14] GENOVESE, C. and WASSERMAN, L. (2004). A stochastic process approach to false discovery control. *Ann. Statist.* **32** 1035–1061. MR2065197

[15] GOEMAN, J. J. and SOLARI, A. (2010). The sequential rejection principle of familywise error control. *Ann. Statist.* **38** 3782–3810. MR2766868

[16] HOLM, S. (1979). A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* **6** 65–70. MR0538597

[17] HOMMEL, G. and HOFFMANN, T. (1988). Controlled uncertainty. In *Multiple Hypothesenprüfung/Multiple Hypotheses Testing* 154–161. Springer, Berlin.

[18] INGLOT, T. (2010). Inequalities for quantiles of the chi-square distribution. *Probab. Math. Statist.* **30** 339–351. MR2792589

[19] INGSTER, Y. I. (1993). Asymptotically minimax hypothesis testing for nonparametric alternatives. I, II, III. *Math. Methods Statist.* **2** 85–114, 171–189, 249–268.

[20] INGSTER, Y. I. (2000). Adaptive chi-square tests. *J. Math. Sci.* **99** 1110–1119.

[21] KORN, E. L., TROENDLE, J. F., MCSHANE, L. M. and SIMON, R. (2004). Controlling the number of false discoveries: Application to high-dimensional genomic data. *J. Statist. Plann. Inference* **124** 379–398. MR2080371

[22] LAURENT, B., LOUBES, J.-M. and MARTEAU, C. (2012). Non asymptotic minimax rates of testing in signal detection with heterogeneous variances. Available at http://arxiv.org/abs/0912.2423v1.

[23] LAURENT, B., LOUBES, J.-M. and MARTEAU, C. (2012). Non asymptotic minimax rates of testing in signal detection with heterogeneous variances. *Electron. J. Stat.* **6** 91–122. MR2879673

[24] LEHMANN, E. L. and ROMANO, J. P. (2005). Generalizations of the familywise error rate. *Ann. Statist.* **33** 1138–1154. MR2195631

[25] LEHMANN, E. L., ROMANO, J. P. and SHAFFER, J. P. (2005). On optimality of stepdown and stepup multiple test procedures. *Ann. Statist.* **33** 1084–1108. MR2195629

[26] MARCUS, R., PERITZ, E. and GABRIEL, K. R. (1976). On closed testing procedures with special reference to ordered analysis of variance. *Biometrika* **63** 655–660. MR0468056

[27] ROMANO, J. P., SHAIKH, A. and WOLF, M. (2011). Consonance and the closure method in multiple testing. *Int. J. Biostat.* **7** 1–27. MR2775079

[28] ROMANO, J. P. and SHAIKH, A. M. (2006). Stepup procedures for control of generalizations of the familywise error rate. *Ann. Statist.* **34** 1850–1873. MR2283720

[29] ROMANO, J. P. and WOLF, M. (2005). Exact and approximate stepdown methods for multiple hypothesis testing. *J. Amer. Statist. Assoc.* **100** 94–108. MR2156821

[30] ROMANO, J. P. and WOLF, M. (2007). Control of generalized error rates in multiple testing. *Ann. Statist.* **35** 1378–1408. MR2351090

[31] ROMANO, J. P. and WOLF, M. (2010). Balanced control of generalized error rates. *Ann. Statist.* **38** 598–633. MR2590052

[32] SCHMIDT-HIEBER, J., MUNK, A. and DÜMBGEN, L. (2013). Multiscale methods for shape constraints in deconvolution: Confidence statements for qualitative features. *Ann. Statist.* **41** 1299–1328. MR3113812

[33] SIMES, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika* **73** 751–754. MR0897872

[34] SPJØTVOLL, E. (1972). On the optimality of some multiple comparison procedures. *Ann. Math. Stat.* **43** 398–411. MR0301871

[35] SPOKOINY, V. G. (1996). Adaptive hypothesis testing using wavelets. *Ann. Statist.* **24** 2477–2498. MR1425962

M. FROMONT
UNIV. RENNES 2
PLACE DU RECTEUR H. LE MOAL
35043 RENNES CEDEX
FRANCE
E-MAIL: magalie.fromont@univ-rennes2.fr

M. LERASLE
P. REYNAUD-BOURET
UNIVERSITÉ DE NICE SOPHIA-ANTIPOLIS
LJAD
06100 NICE
FRANCE
E-MAIL: mlerasle@unice.fr
reynaudb@unice.fr