Submitted to Bernoulli

Consistency of p-norm based tests in high dimensions: characterization, monotonicity, domination

ANDERS BREDAHL KOCK[1] and DAVID PREINERSTORFER[2]

[1]University of Oxford, Aarhus University and CREATES, 10 Manor Rd, Oxford OX1 3UQ, anders.kock@economics.ox.ac.uk
[2]University of St.Gallen, SEPS-SEW, Varnbüelstrasse 14, 9000 St.Gallen, david.preinerstorfer@unisg.ch

Many commonly used test statistics are based on a norm measuring the evidence against the null hypothesis. To understand how the choice of that norm affects power properties of tests in high dimensions, we study the *consistency sets* of p-norm based tests in the prototypical framework of sequence models with unrestricted parameter spaces, the null hypothesis being that all observations have zero mean. The consistency set of a test is here defined as the set of all arrays of alternatives the test is consistent against as the dimension of the parameter space diverges. We characterize the consistency sets of p-norm based tests and find, in particular, that the consistency against an array of alternatives cannot be determined solely in terms of the p-norm of the alternative. Our characterization also reveals an unexpected monotonicity result: namely that the consistency set is strictly increasing in $p \in (0, \infty)$, such that tests based on higher p strictly dominate those based on lower p in terms of consistency. This monotonicity property allows us to construct novel tests that dominate, with respect to their consistency behavior, *all* p-norm based tests without sacrificing asymptotic size.

Keywords: Consistency; high-dimensional testing problems; norm-based tests; power enhancement principle; sequence models

1. Introduction

Since the advent of the Big Data era, the development of procedures for testing hypotheses in high-dimensional models has received considerable attention with applications in fields such as genomics, finance, economics and engineering. Many tests are based on a p-norm, measuring the evidence against the null hypothesis. In contrast to classical low-dimensional testing problems, tests based on different p-norms tend to be consistent against different types of alternatives in the high-dimensional setting, cf., e.g., Ingster and Suslina (2003). This observation led Fan, Liao and Yao (2015) to put forward their remarkable power enhancement principle, a combination procedure that improves an initial test by combining it with another one. The resulting *enhanced* test has the same asymptotic size as the initial test but is consistent against more alternatives. Specifically, Fan, Liao and Yao (2015) enhanced a Euclidean norm based test by combining it with a supremum-norm based test, thus constructing a test that: (i) retains the good consistency properties of the Euclidean norm based test against dense alternatives and (ii) is also consistent against a large class of sparse alternatives. This possibility of increasing power by combining tests has led to an intense recent interest in constructions combining tests based on various p-norms in many types of high-dimensional testing problems, cf. Xu et al. (2016), Yang and Pan (2017), Yu, Li and Xue (2020), He et al. (2021), Yu et al. (2021) [testing high-dimensional means and covariance matrices]; Zhang, Wang and Shao (2021) [change point detection]; Jammalamadaka, Meintanis and Verdebout (2020) [tests for uniformity on the sphere]; Feng et al. (2022) [tests for cross-sectional independence in high-dimensional panel data models]. Furthermore, Wang, McKeague and Qian (2018) pointed to the potential usefulness of combining tests based on the Euclidean and supremum-norm in the context of high-dimensional quantile regression. By combining either tests based on the Euclidean and supremum-norm (sometimes referred to as "max-sum" tests) or a finite number of norms, these papers establish that one can construct tests with better power properties compared to tests based on a single p-norm in the concrete testing problems considered.

The classical results in Ingster and Suslina (2003) as well as the more recent contributions in the context of the power enhancement principle of Fan, Liao and Yao (2015) discussed above have in common that *sufficient* conditions for consistency (or inconsistency) of the tests under consideration are established. However, even for a test based on a single p-norm it is not known to date what the necessary and sufficient condition for consistency of such a test is in high dimensions. As a consequence, also the following questions are open:

- Can one obtain tests that are consistent against more alternatives by using p-norms other than the 1-, Euclidean or supremum-norm, e.g., the 4-norm?
- Is it possible to rank p-norm based tests in terms of their consistency properties?

Similarly, in the context of combinations of p-norm based tests, one may ask questions such as:

- When combining several p-norm based tests: how many and which should one combine if one wants to obtain a test with optimal consistency properties?
- Is there a test that simultaneously dominates all p-norm based tests in terms of consistency?

1.1. Contributions

We study the questions raised above in the prototypical context of a sequence model. To reject the "global null" of all observations having zero mean, these tests gauge whether the p-norm of the observation vector exceeds a given critical value. Our contributions are as follows:

- *Characterization*: We characterize the *consistency set* of each p-norm based test, that is the set of arrays of alternatives that the test is consistent against. Somewhat surprisingly, we find that whether the p-norm based test is consistent against an array of alternatives cannot be determined solely by the p-norm of the alternative (which would in general only lead to sufficient but not necessary conditions for consistency). For $p \in (2, \infty)$ it actually suffices that the p-norm *or* the Euclidean norm of the alternative is sufficiently large.
- *Monotonicity*: For $0 < p < q < \infty$ the q-norm based test is consistent against any array of alternative that the p-norm based test is consistent against — and more. This finding sheds light on the consequences of employing tests based on the commonly used 1- and Euclidean norms, as these are strictly dominated in terms of the magnitude of their consistency sets. Note that the strict ranking of p-norm based tests in terms of the magnitude of their consistency sets is in contrast to the absence of such a strict ranking in terms of local power properties, e.g., Pinelis (2010) and Pinelis (2014). We also identify the structure of the additional alternatives that a test based on a larger power is consistent against. This strict dominance relation does not extend to $p = \infty$.
- *Domination*: Although no single p-norm based test dominates all other ones in terms of the magnitude of its consistency set, we construct a single test that is consistent against any array of alternatives that any p-norm based test is consistent against — and more. Thus, it is possible to *simultaneously* dominate all members of the commonly used class of p-norm based tests. Finally, we revisit results given in Ingster and Suslina (2003) and we also establish some new adaptation results in the minimax framework building on our results obtained in earlier sections.

An immediate consequence of our findings for the current practice of either using a Euclidean norm based test, a supremum-norm based test or a combination of the two, is that a test that is consistent against strictly more alternatives than any of these can be explicitly constructed.

2. Framework

2.1. Model and testing problem

We consider a sequence model

$$y_{i,d} = \theta_{i,d} + \varepsilon_i, \quad i = 1, \ldots, d, \tag{1}$$

where the $\theta_{i,d} \in \mathbb{R}$ are unknown parameters, the unobserved error terms ε_i have mean zero and are i.i.d., and where $y_{1,d}, \ldots, y_{d,d}$ are the observations. *For simplicity of presentation, and because it is the most important case, we shall assume throughout the main text of this article that* $\varepsilon_i \sim \mathbb{N}(0, 1)$. All results generalize appropriately to non-normal errors under suitable assumptions concerning the tail of the distribution of ε_i. Such results are established in Appendix C in the Supplementary Material (Kock and Preinerstorfer (2021)).

For notational simplicity, write $\boldsymbol{y}_d = (y_{1,d}, \ldots y_{d,d})$, $\boldsymbol{\varepsilon}_d = (\varepsilon_1, \ldots, \varepsilon_d)$, and $\boldsymbol{\theta}_d = (\theta_{1,d}, \ldots, \theta_{d,d})$, upon which we may write (1) equivalently as $\boldsymbol{y}_d = \boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d$.

In the sequence model (1), we study properties of tests for the testing problem

$$H_{0,d} : \boldsymbol{\theta}_d = \mathbf{0}_d \quad \text{against} \quad H_{1,d} : \boldsymbol{\theta}_d \in \mathbb{R}^d \setminus \{\mathbf{0}_d\}, \tag{2}$$

where $\mathbf{0}_d = (0, \ldots, 0) \in \mathbb{R}^d$. That is, we are concerned with what is sometimes referred to as the "global null" of no effect. Oftentimes, a test for this hypothesis is done in the initial stage of a whole series of investigations that are carried out only if the global null is rejected. It is then crucial to use a test with good power properties at the initial stage.

The sequence model (1) is an often used prototypical framework in high-dimensional statistics. On the one hand, the simplicity of the model allows one to focus on asymptotic power properties as the dimension d diverges to ∞ in a clean framework. On the other hand, the framework assumes away certain complications, such as the estimation of nuisance (e.g., variance) parameters, or the interplay between sample size and the dimensionality of the parameter vector, which we do not address.

The following two remarks illustrate how testing problems in the Gaussian sequence model are related to those in other statistical models.

Remark 2.1. Let $\boldsymbol{X}_d \in \mathbb{R}^{n\times d}$ be a (non-random) matrix of full column rank, in particular implying $d \leq n$. In the Gaussian linear regression model $\mathbf{z}_d = \boldsymbol{X}_d\boldsymbol{\beta}_d + \mathbf{u}_d$ with $\mathbf{u}_d \sim \mathbb{N}(\mathbf{0}_d, \boldsymbol{I}_d)$, the OLS estimator $\hat{\boldsymbol{\beta}}_d$ is sufficient for $\boldsymbol{\beta}_d$ and, denoting $\boldsymbol{M}_d := (\boldsymbol{X}'_d\boldsymbol{X}_d)^{1/2}$, the same holds for $\boldsymbol{M}_d\hat{\boldsymbol{\beta}}_d$, which satisfies

$$\boldsymbol{M}_d\hat{\boldsymbol{\beta}}_d := \boldsymbol{M}_d\boldsymbol{\beta}_d + \boldsymbol{M}_d^{-1}\boldsymbol{X}'_d\mathbf{u}_d \quad \text{with} \quad \boldsymbol{M}_d^{-1}\boldsymbol{X}'_d\mathbf{u}_d \sim \mathbb{N}(\mathbf{0}_d, \boldsymbol{I}_d),$$

which is clearly of the form (1). Note also that $\boldsymbol{\theta}_d := \boldsymbol{M}_d\boldsymbol{\beta}_d = \mathbf{0}_d$ if and only if $\boldsymbol{\beta}_d = \mathbf{0}_d$. That is, this problem is statistically equivalent to testing (2) in a (re-parameterized) Gaussian sequence model.

Remark 2.2. On a *conceptual and informal* level, results obtained in Gaussian sequence models carry over much more broadly to any situation where the distribution of a properly standardized estimator is approximately Gaussian. The reasoning follows along similar lines as in Remark 2.1: Consider a situation where an estimator $\hat{\boldsymbol{\beta}}_d$ for a target parameter $\boldsymbol{\beta}_d \in \mathbb{R}^d$ is available the distribution of which satisfies

$$\hat{\boldsymbol{\beta}}_d \approx \mathbb{N}(\boldsymbol{\beta}_d, \boldsymbol{\Omega}_d).$$

Suppose further that an invertible estimator $\hat{\boldsymbol{\Omega}}_d \approx \boldsymbol{\Omega}_d$ is available, such that

$$\hat{\boldsymbol{\Omega}}_d^{-1/2}\hat{\boldsymbol{\beta}}_d \approx \mathbb{N}(\boldsymbol{\Omega}_d^{-1/2}\boldsymbol{\beta}_d, \boldsymbol{I}_d).$$

Then, testing $\boldsymbol{\beta}_d = \boldsymbol{0}_d$ on the basis of $\hat{\boldsymbol{\beta}}_d$ and $\hat{\boldsymbol{\Omega}}_d$ is approximated by testing $\boldsymbol{\theta}_d := \boldsymbol{\Omega}_d^{-1/2}\boldsymbol{\beta}_d = \boldsymbol{0}_d$ in a Gaussian sequence model. Precise sets of conditions under which the above approximation statements hold depend on the interplay of the dimension of the target parameter and sample size and on particularities of the specific setup under consideration.

2.2. Tests, asymptotic size, and consistency sets

Suppose that for every $d \in \mathbb{N}$ we are given a (possibly randomized) test φ_d for the testing problem (2). That is, φ_d is a (Borel measurable) function from $\mathbb{R}^d$ to $[0, 1]$. We denote the set of all such sequences of tests $\{\varphi_d\}_{d\in\mathbb{N}}$ by $\mathbb{T}$. With some abuse of notation, we shall often abbreviate $\{\varphi_d\}_{d\in\mathbb{N}}$ by φ_d, to which we then also refer to as a sequence of tests or just as a test.

We say that the sequence of tests $\varphi_d \in \mathbb{T}$ has *asymptotic size* $\alpha \in [0, 1]$ if and only if

$$\lim_{d\to\infty} \mathbb{E}\left(\varphi_d(\boldsymbol{\varepsilon}_d)\right) = \alpha. \tag{3}$$

The subset of all sequences of tests with asymptotic size α is denoted by $\mathbb{T}_\alpha \subseteq \mathbb{T}$. We will mostly be concerned with $\alpha \in (0, 1)$.

We consider an asymptotic framework where the coordinates of $\boldsymbol{\theta}_d = (\theta_{1,d}, \ldots, \theta_{d,d})$ depend on d. That is, we consider (uniform) power properties along triangular arrays of alternatives $\boldsymbol{\vartheta} = \{\boldsymbol{\theta}_d : d \in \mathbb{N}\}$, where $\boldsymbol{\theta}_d \in \mathbb{R}^d$ for every $d \in \mathbb{N}$. The set of all such triangular arrays will be denoted by $\boldsymbol{\Theta} := \bigtimes_{d=1}^{\infty} \mathbb{R}^d$. As is common, we say that the sequence of tests φ_d is *consistent* against an array $\boldsymbol{\vartheta} = \{\boldsymbol{\theta}_d : d \in \mathbb{N}\} \in \boldsymbol{\Theta}$ if and only if

$$\lim_{d\to\infty} \mathbb{E}\left(\varphi_d(\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d)\right) = 1. \tag{4}$$

To any sequence of tests $\varphi_d \in \mathbb{T}$ we associate its *consistency set* $\mathscr{C}(\varphi_d) \subseteq \boldsymbol{\Theta}$, i.e., the subset of all arrays of alternatives $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$ that the sequence of tests φ_d is consistent against. The characterization and comparison of consistency sets is the main focus of the present article.

2.3. Optimal consistency properties

For a given $\alpha \in (0, 1)$, one can compare the quality of two sequences of tests $\psi_d \in \mathbb{T}_\alpha$ and $\varphi_d \in \mathbb{T}_\alpha$ on the basis of their consistency sets.[1] By definition, $\mathscr{C}(\psi_d) \supseteq \mathscr{C}(\varphi_d)$ if and only if ψ_d is consistent against *all* alternatives that φ_d is consistent against. If this is the case, ψ_d weakly dominates φ_d in terms of consistency, whereas ψ_d strongly dominates φ_d in terms of consistency if even $\mathscr{C}(\psi_d) \supsetneq \mathscr{C}(\varphi_d)$.

---

[1] In principle, one could also compare two sequences of tests with different asymptotic sizes by comparing their consistency sets, but the comparison is less meaningful in such situations.

In search for an "optimal" test, one would hope that there exists a sequence of tests $\psi_d \in \mathbb{T}_\alpha$ that is consistent against all alternatives that some $\varphi_d \in \mathbb{T}_\alpha$ is consistent against, i.e., ψ_d satisfies

$$\mathscr{C}(\psi_d) \supseteq \bigcup \{\mathscr{C}(\varphi_d) : \varphi_d \in \mathbb{T}_\alpha\}. \tag{5}$$

Such a test ψ_d (in case it exists) would be "consistency-optimal" in the sense that no other test with the same asymptotic size exists that is consistent against more alternatives.

Whereas the optimality property in (5) is often achieved by the usual candidates of tests in standard finite-dimensional testing problems, the consistency properties of tests are much more delicate in high-dimensional testing problems, where a $\psi_d \in \mathbb{T}_\alpha$ satisfying (5) typically does *not* exist. This is illustrated in the following result, which can be obtained by a similar reasoning as in Section 1.1 of Kock and Preinerstorfer (2019), to which we refer for further discussions and results in the context of high-dimensional models that are locally asymptotically normal in a suitable sense.

Theorem 2.1. *Let $\alpha \in (0,1)$. For every $\psi_d \in \mathbb{T}_\alpha$ there exists a $\overline{\psi}_d \in \mathbb{T}_\alpha$, such that:*

1. *$\overline{\psi}_d \geq \psi_d$ holds for every $d \in \mathbb{N}$, guaranteeing that $\overline{\psi}_d$ has uniformly non-inferior asymptotic power compared to ψ_d;*
2. *$\overline{\psi}_d$ is consistent against an array of alternatives against which ψ_d has asymptotic power at most α.*

In particular it holds that $\mathscr{C}(\psi_d) \subsetneqq \mathscr{C}(\overline{\psi}_d)$.

Theorem 2.1 establishes that the consistency set of *any* test (of asymptotic size $\alpha \in (0,1)$) can be strictly improved.[2] Inspection of the proof, which largely builds on an argument in Section 3.4.2 of Ingster and Suslina (2003), shows that the improvement can be achieved through the power enhancement principle of Fan, Liao and Yao (2015). Hence, no consistency-optimal test as fancied in the second paragraph of this section and (5) in particular can exist.

The literature mainly offers two solutions in situations where (for a given optimality criterion) no optimal test exists: The first is to restrict one's attention to a certain subclass of alternatives. Restricting the class of alternatives is an approach that is often employed in high-dimensional and nonparametric problems, e.g., by focusing on sparse alternatives or all alternatives for which a given norm exceeds a certain threshold. One then explores and constructs tests that work well for such alternatives (but may perhaps not work so well for others). This approach then typically studies minimax rates of detection, and minimax consistent tests. We offer more discussion and provide some results in that direction in Section 6.

A different approach is to leave the set of alternatives unrestricted as is, but instead focus on a specific class of tests, and to study optimality for this restricted class of tests. In essence, and focusing on consistency properties, one fixes a subset $\mathbb{T}^*_\alpha \subseteq \mathbb{T}_\alpha$ and looks for a test whose consistency set contains the consistency sets of all tests in $\mathbb{T}^*_\alpha$; that is one searches for a $\psi_d \in \mathbb{T}_\alpha$ that satisfies (5) but with $\mathbb{T}_\alpha$ replaced by $\mathbb{T}^*_\alpha$. Whether or not a test ψ_d with the desired property exists clearly depends on $\mathbb{T}^*_\alpha$. In the present article we explore this question for the important subclass of p-norm based tests.

---

[2] Actually one can show that *any* test, including those based on the supremum-norm, can be strictly improved against a highly sparse array of alternatives; cf. Theorem A.1 in Appendix A. In this sense, not even the most stringent sparsity assumptions on the arrays of alternatives considered overturn the conclusion of Theorem 2.1.

2.4. p-norm based tests

For $\boldsymbol{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$ and $p \in (0, \infty]$, we define

$$\|\boldsymbol{x}\|_p := \begin{cases} \left(\sum_{i=1}^{d} |x_i|^p \right)^{\frac{1}{p}} & \text{if } p < \infty, \\ \max_{i=1,\dots,d} |x_i| & \text{else.} \end{cases} \tag{6}$$

For $p \geq 1$ the function $\| \cdot \|_p$ defines a norm on $\mathbb{R}^d$; whereas for $p \in (0, 1)$ it only defines a quasinorm on $\mathbb{R}^d$. For the sake of brevity, we will refer to $\| \cdot \|_p$ as the "p-norm" also in case $p \in (0, 1)$. Given a radius $r \in [0, \infty]$, we denote $\mathbb{B}_p^d(r) := \{\boldsymbol{y} \in \mathbb{R}^d : \|\boldsymbol{y}\|_p \leq r\}$, i.e., the closed "ball" with respect to $\| \cdot \|_p$ of radius r centered at the origin.

It is well known that the likelihood ratio test for (2) rejects if $\|\boldsymbol{y}_d\|_2$, the Euclidean norm of the vector of observations $\boldsymbol{y}_d$, exceeds a critical value (chosen to satisfy a given size constraint). More generally, any p-norm delivers a test for (2), i.e., one rejects the null hypothesis if $\|\boldsymbol{y}_d\|_p$ exceeds some critical value. Given $p \in (0, \infty]$ and a sequence of critical values κ_d, we abbreviate the sequence of tests $\mathbb{1}\{\| \cdot \|_p \geq \kappa_d\}$ by $\{p, \kappa_d\}$, and refer to such a test as a *p-norm based* test. Consequently, the consistency set of such a sequence of tests is written as $\mathscr{C}(\{p, \kappa_d\})$.

By a classical result of Birnbaum (1955) and Stein (1956), tests that reject if the p-norm of $\boldsymbol{y}_d$ exceeds a given critical value are *admissible* for every $d \in \mathbb{N}$ and every $p \in [1, \infty]$; the reason is that their acceptance region is convex. Furthermore, the tests just described are all *unbiased*, for every $d \in \mathbb{N}$, due to Anderson's theorem, cf. Anderson (1955). Hence, all p-norm based tests with $p \in [1, \infty]$ are reasonable from a non-asymptotic point of view.

Following the Neyman-Pearson approach, we shall mostly be interested in the situation where the critical values κ_d are chosen such that the asymptotic size of $\{p, \kappa_d\}$ is in $(0, 1)$, i.e.,

$$\lim_{d \to \infty} \mathbb{P}\left(\|\boldsymbol{\varepsilon}_d\|_p \geq \kappa_d\right) = \alpha \in (0, 1). \tag{7}$$

One has that (7) is true if and only if

$$\kappa_d = \begin{cases} \left[\left(\Phi^{-1}(1 - \alpha) + o(1) \right) \sqrt{d} \sigma_p + d\mu_p \right]^{1/p} & \text{if } p \in (0, \infty), \\ \sqrt{2 \log(d)} - \dfrac{\log\log(d) + \log(4\pi) + 2\log(-\log(1 - \alpha)/2) + o(1)}{2\sqrt{2\log(d)}} & \text{if } p = \infty. \end{cases} \tag{8}$$

Here Φ denotes the cdf of the standard normal distribution and for $p \in (0, \infty)$ we abbreviated $\sigma_p^2 := \mathbb{V}ar(|\varepsilon_1|^p)$ and $\mu_p := \mathbb{E}(|\varepsilon_1|^p)$. The equality in (8) follows since for $p \in (0, \infty)$ the test statistic converges in distribution under the null (upon suitable centering and scaling) to the standard normal distribution, whereas for $p = \infty$ it converges to a slight modification (only taking care of absolute values) of Gumbel's double exponential distribution (cf. Lemma C.3 and Lemma A.2 for formal statements).

3. Consistency sets of p-norm based tests

We shall now derive a characterization of $\mathscr{C}(\{p, \kappa_d\})$ for sequences κ_d satisfying (7). This characterization will in particular allow us to compare the consistency sets for different $p \in (0, \infty]$. The present section is divided into the cases $p \in (0, \infty)$ and $p = \infty$, which lead to fundamentally different results.

Characterizing the consistency set of a p-norm based test requires us to establish a necessary and sufficient condition for the test to be consistent against an array of alternatives. For p-norm based tests a set of sufficient conditions is also given in Ingster and Suslina (2003), Sections 3.1.2–3.1.4. However, these conditions are not necessary. That is, there are arrays $\boldsymbol{\vartheta}$ against which a p-norm based test is consistent, but which do not satisfy the sufficient condition for consistency there.

3.1. Case $p \in (0, \infty)$

For any $p \in \mathbb{R}$, denote by $g_p : \mathbb{R} \to \mathbb{R}$ the function

$$g_p(x) := x^2 \mathbb{1}_{[-1,1]}(x) + |x|^p \mathbb{1}_{\mathbb{R} \setminus [-1,1]}(x). \tag{9}$$

In particular, $g_p(x) = x^2 \wedge |x|^p$ for $p \in (0, 2)$ and $g_p(x) = x^2 \vee |x|^p$ for $p \in [2, \infty)$, where for two real numbers x and y we write $\min(x, y) = x \wedge y$ and $\max(x, y) = x \vee y$ for readability.

The consistency set of a p-norm based test with non-trivial asymptotic size is characterized next.

Theorem 3.1. *For $p \in (0, \infty)$ and $\{p, \kappa_d\} \in \mathbb{T}_\alpha$, $\alpha \in (0, 1)$, we have*

$$\boldsymbol{\vartheta} \in \mathscr{C}(\{p, \kappa_d\}) \quad \Leftrightarrow \quad \frac{\sum_{i=1}^d g_p(\theta_{i,d})}{\sqrt{d}} \to \infty. \tag{10}$$

It is a somewhat surprising aspect of Theorem 3.1 that, apart from the case $p = 2$, the consistency set $\mathscr{C}(\{p, \kappa_d\})$ cannot be entirely characterized in terms of the asymptotic behavior of the p-norm of the elements of an array $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$. Regardless of p, coordinates of $\boldsymbol{\theta}_d$ that are small in absolute value enter the consistency criterion in (10) via their squares, whereas coordinates with large absolute values enter differently and in dependence on p. We illustrate this in Figure 1 by showing contour plots of the level sets of the function $(x_1, x_2) \mapsto g_p(x_1)/\sqrt{2} + g_p(x_2)\sqrt{2}$ for $p = 1, 3$, which are genuinely different from the level sets of the corresponding p-norm in dimension $d = 2$. Note that the level sets close to the origin are circular irrespective of the value of p, i.e., in correspondence to level sets of a Euclidean norm, whereas level sets further away from the origin approach those of the p-norm the test is based on.

Theorem 3.1 follows from a more general result that also holds in non-Gaussian settings and which is given in Theorem C.4 of Appendix C in the Supplementary Material (Kock and Preinerstorfer (2021)). All results in the present Section 3.1 carry over to this more general setting, cf. the discussion after the proof of Theorem C.4.

Remark 3.1. There is nothing special about the interval $[-1, 1]$ in the definition of g_p in Equation (9). In principle, one could replace the interval $[-1, 1]$ in the definition of g_p by any interval $[-M, M]$, for a fixed $M > 0$, and the statement in (10) would still be correct with this re-defined function $g_p^{(M)}$, say, in place of g_p. This follows immediately from

$$0 < \inf_{x \in \mathbb{R} \setminus \{0\}} \frac{g_p^{(M)}(x)}{g_p(x)} \leq \sup_{x \in \mathbb{R} \setminus \{0\}} \frac{g_p^{(M)}(x)}{g_p(x)} < \infty.$$

We have chosen $M = 1$ in the formulation of Theorem 3.1 for concreteness and because it is the most convenient choice for later use in the proof of Theorem 3.3. The content of this remark will be instrumental in establishing Theorem 3.4 further below.

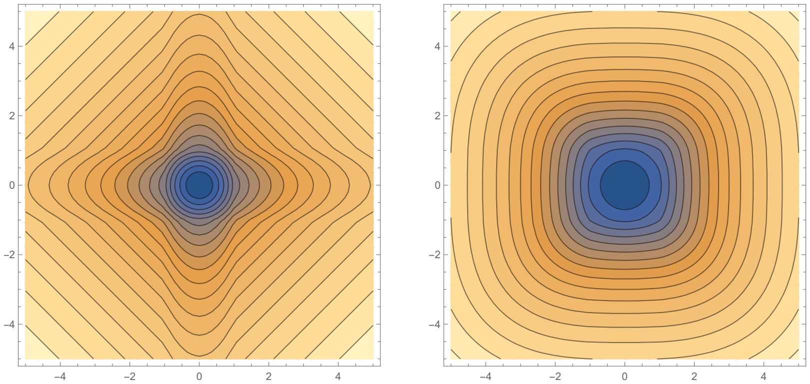

Figure 1: Contour plots of the function characterizing the consistency of the p-norm based test for $d = 2$ and $p = 1$ (left panel) and $p = 3$ (right panel).

We emphasize (also for later use in Section 5) that the condition on the right in Equation (10) does not depend on α. That is, as long as the sequence of critical values κ_d is chosen such that the asymptotic size of the corresponding test equals $\alpha \in (0,1)$ (cf. also (8)), the consistency set does not depend on the concrete value of α.

An immediate consequence of Theorem 3.1 is the following observation, which we shall make use of later. In particular, it contains an equivalent way of writing (10) in case $p \in [2,\infty)$ which further clarifies that the consistency of the p-norm based test against any array of alternatives cannot be settled based on the sequence $\|\boldsymbol{\theta}_d\|_p$ alone. The condition $\|\boldsymbol{\theta}_d\|_p^p/\sqrt{d} \to \infty$, which is also given in Ingster and Suslina (2003) Corollary 3.6, is only sufficient.

Corollary 3.2. *For $p \in (0,\infty)$ and $\{p,\kappa_d\} \in \mathbb{T}_\alpha$, $\alpha \in (0,1)$, we have:*

1. *If $p \in [2,\infty)$, then*

$$\boldsymbol{\vartheta} \in \mathscr{C}(\{p,\kappa_d\}) \quad \Leftrightarrow \quad d^{-1/2}\left(\|\boldsymbol{\theta}_d\|_2^2 \vee \|\boldsymbol{\theta}_d\|_p^p\right) \to \infty. \tag{11}$$

2. *If $p \in (0,2)$,*

$$\boldsymbol{\vartheta} \in \mathscr{C}(\{p,\kappa_d\}) \quad \Rightarrow \quad d^{-1/2}\left(\|\boldsymbol{\theta}_d\|_2^2 \wedge \|\boldsymbol{\theta}_d\|_p^p\right) \to \infty; \tag{12}$$

   but the condition to the right in (12) does not imply the condition to the left.

We shall now highlight a “monotonicity” result concerning the consistency set of p-norm based tests, which can be obtained from Theorem 3.1. This result shows that in terms of consistency properties, and choosing among p-norm based tests with $p \in (0,\infty)$ and asymptotic size in $(0,1)$, it is best to choose p “large.” [3]

---

[3] Because the statement of Theorem 3.3 involves tests based on different exponents and the sequences of critical values effecting asymptotic size control depend on the respective exponent, we here explicitly indicate this dependence simply to distinguish the two sequences. A similar convention will be applied in the statements below when needed.

Theorem 3.3. *For sequences of tests* $\{p, \kappa_{d,p}\}$ *and* $\{q, \kappa_{d,q}\}$ *with asymptotic sizes in* $(0,1)$*, and* $0 < p < q < \infty$*, we have*

$$\mathscr{C}(\{p, \kappa_{d,p}\}) \subsetneqq \mathscr{C}(\{q, \kappa_{d,q}\}).$$

In other words, the larger $p \in (0,\infty)$, the larger the consistency set of $\{p, \kappa_{d,p}\}$. Somewhat unfortunately, however, the theorem shows — at the same time — that it is impossible to choose p "large enough." That tests based on larger powers have larger consistency sets may appear contradictory in light of the inequality $\|\cdot\|_p \geq \|\cdot\|_q$ for $1 \leq p \leq q$, which implies that the test statistic gets *smaller* for large powers. Note, however, that this is compensated by the critical values, cf. (8). We finally emphasize that the asymptotic sizes of the two sequences of tests in Theorem 3.3 do not need to be identical. This is a consequence of the independence of the consistency set of $\alpha \in (0,1)$, cf. Theorem 3.1.

Returning to the discussion in Section 2.3, the monotonicity statement established in Theorem 3.3 immediately raises the question whether there exists a test of asymptotic size $\alpha \in (0,1)$ that is consistent against all sequences of alternatives that any p-norm based test is consistent against; that is:

Question 1. Given $\alpha \in (0,1)$, does there exist a $\psi_d \in \mathbb{T}_\alpha$ such that

$$\mathscr{C}(\{p, \kappa_{d,p}\}) \subseteq \mathscr{C}(\psi_d) \tag{13}$$

for every $p \in (0,\infty)$ and every sequence $\kappa_{d,p}$ such that $\{p, \kappa_{d,p}\} \in \mathbb{T}_\alpha$?

Remark 3.2. It follows immediately from Theorem 3.3 that the requirement on ψ_d in (13) of Question 1 can equivalently be replaced by

$$\mathscr{C}(\{p, \kappa_{d,p}\}) \subsetneqq \mathscr{C}(\psi_d); \tag{14}$$

that is, if a test as in Question 1 exists, then it must not only contain the consistency set of every p-norm based test ($p \in (0,\infty)$)) as a subset, but this set inclusion must actually be strict for every p. In this sense, if a test ψ_d as in Question 1 exists, it *strictly* dominates all p-norm based tests in terms of their consistency behavior.

We will come back to Question 1 in later sections. There, it will also be shown that despite the monotonicity established in Theorem 3.3 for $p \in (0,\infty)$, the supremum-norm based (i.e., $p = \infty$) test does not have the property of ψ_d sought in Question 1.

In the context of Theorem 3.3 one may ask whether the arrays of alternatives that the q-norm based test but not the p-norm based test is consistent against have particular structural properties. Thus, we next study the structure of

$$\mathscr{C}(\{q, \kappa_{d,q}\}) \setminus \mathscr{C}(\{p, \kappa_{d,p}\}) \qquad \text{for } p < q. \tag{15}$$

3.1.1. Structure of elements of (15)

Let $2 \leq p < q < \infty$ and fix critical values such that $\{p, \kappa_{d,p}\}$ and $\{q, \kappa_{d,q}\}$ have asymptotic sizes in $(0,1)$. Then Corollary 3.2 shows that $\boldsymbol{\vartheta}$ is an element of the set in (15) if and only if (i) there exists a subsequence along which $d^{-1/2}(\|\boldsymbol{\theta}_d\|_2^2 \vee \|\boldsymbol{\theta}_d\|_p^p)$ converges to a real number and (ii) if along any such subsequence it holds that $d^{-1/2}\|\boldsymbol{\theta}_d\|_q^q \to \infty$. A concrete example of such an array is

$$\boldsymbol{\vartheta} = \{(d^{\frac{1}{2p}}, 0, \ldots, 0) : d \in \mathbb{N}\},$$

i.e., sparse alternatives with a dominating coordinate diverging to ∞ at an appropriate rate. That all elements $\boldsymbol{\vartheta}$ in the set in (15) are approximately sparse and highly unbalanced, at least along suitably chosen subsequences, will be shown next; the proof makes use of the observation in Remark 3.1.

Theorem 3.4. *Let $0 < p < q < \infty$, and let $\{p, \kappa_{d,p}\}$ and $\{q, \kappa_{d,q}\}$ have asymptotic sizes in $(0,1)$. Then, for every $\boldsymbol{\vartheta} \in \mathscr{C}(\{q, \kappa_{d,q}\}) \setminus \mathscr{C}(\{p, \kappa_{d,p}\})$ the following holds:*

1. *For every $\delta \in (0, \infty)$ there exists a subsequence d' of d, such that*

$$\sup_{d'} \frac{\sum_{i=1}^{d'} \mathbb{1}\{|\theta_{i,d'}| > \delta\}}{\sqrt{d'}} < \infty \quad \text{and} \quad \max_{1 \le i \le d'} |\theta_{i,d'}| \to \infty. \tag{16}$$

2. *If $p \geq 2$, then there exists a subsequence d' of d, such that*

$$\sup_{\delta \in (0,\infty)} \delta^p \times \sup_{d'} \frac{\sum_{i=1}^{d'} \mathbb{1}\{|\theta_{i,d'}| > \delta\}}{\sqrt{d'}} < \infty \quad \text{and} \quad \max_{1 \le i \le d'} |\theta_{i,d'}| \to \infty. \tag{17}$$

The first part of Theorem 3.4 shows that those arrays $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$ that are consistently detected by some p-norm based test, but only if p is chosen large enough, are, along a subsequence, (i) "approximately sparse," in that they have a vanishing fraction of entries larger than a given $\delta > 0$, and (ii) "highly unbalanced," in the sense that

$$\min_{1 \le i \le d} |\theta_{i,d}| / \max_{1 \le i \le d} |\theta_{i,d}| \approx 0.$$

Furthermore, the second part of Theorem 3.4 shows that in case $\boldsymbol{\vartheta}$ is not consistently detected by a p-norm based test with $p \geq 2$ (but is detected by some q-norm based test with $q > p$), a bit more can be said concerning (i). Namely that then the fraction of entries larger than a given $\delta > 0$ decays (at least) at the order δ^{-p}, uniformly over a *common* subsequence d'.

3.2. Case $p = \infty$

We shall now present a result that characterizes the consistency set of tests $\{\infty, \kappa_d\} \in \mathbb{T}_\alpha$, $\alpha \in (0,1)$. Structurally, the result is comparable to the one given in Theorem 3.1. Define

$$g_\infty(x) := \begin{cases} w(x) & \text{if } x < 1 \\ e^{-x^2/2}/x & \text{else,} \end{cases} \tag{18}$$

for a fixed continuous function $w : \mathbb{R} \to (0, \infty)$ that satisfies $w(z) \to \infty$ as $z \to -\infty$; e.g., $w(z) = e^{(z^2 - 2z)/2}$, in which case g_∞ is also continuous and strictly decreasing. The result is as follows, where we recall that we denote the standard Gaussian cdf by Φ and set $\overline{\Phi} := 1 - \Phi$. Its proof is based on a result applicable beyond the Gaussian setting which can be found in Proposition C.6 in Appendix C.2 in the Supplementary Material (Kock and Preinerstorfer (2021)).

Theorem 3.5. *For κ_d such that $\{\infty, \kappa_d\} \in \mathbb{T}_\alpha$, $\alpha \in (0,1)$, we have*

$$\boldsymbol{\vartheta} \in \mathscr{C}(\{\infty, \kappa_d\}) \Leftrightarrow \sum_{i=1}^{d} \frac{\overline{\Phi}\left(\mathfrak{c}_d - |\theta_{i,d}|\right)}{\Phi\left(\mathfrak{c}_d - |\theta_{i,d}|\right)} \to \infty \Leftrightarrow \sum_{i=1}^{d} g_\infty\left(\mathfrak{c}_d - |\theta_{i,d}|\right) \to \infty, \tag{19}$$

where $\mathfrak{c}_d := \sqrt{2\log(d)} - \frac{\log\log(d)}{2\sqrt{2\log(d)}}$ for $d \geq 2$ (and one may set $\mathfrak{c}_1 := 0$ for completeness).

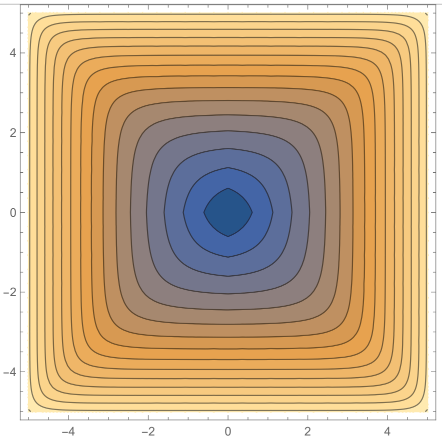

Figure 2: Contour plots of the function characterizing the consistency of the supremum-norm based test for $d = 2$.

First of all, we note that the sequence $\mathfrak{c}_d$ in (19) plays the role of centering. This is in contrast to (10), where the function g_p is symmetric about 0 and standardization is achieved by multiplying by $1/\sqrt{d}$.

In analogy to the discussion after Theorem 3.1 (cf. also Figure 1), we observe that Theorem 3.5 shows that the consistency set of a supremum-norm based test is not characterized exclusively by the supremum-norm of the deviation from the null hypothesis. Figure 2 further clarifies this by showing the contour sets for the function $(x_1, x_2) \mapsto \overline{\Phi}(\mathfrak{c}_2 - |x_1|)/\Phi(\mathfrak{c}_2 - |x_1|) + \overline{\Phi}(\mathfrak{c}_2 - |x_2|)/\Phi(\mathfrak{c}_2 - |x_2|)$. We note that the contour sets deviate from the ones of a supremum-norm, in particular along the "diagonals" or close to the origin.

Theorem 3.5 also reveals that as long as the critical values κ_d are chosen such that the asymptotic size of a supremum-norm based test is in $(0, 1)$, the consistency set remains the same, i.e., does not depend on the concrete asymptotic size. Although this already follows from the first equivalence in Theorem 3.5, the second equivalence statement has the advantage that it is easier to use in order to check whether a particular $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$ is in $\mathscr{C}(\{\infty, \kappa_d\})$.

Concerning "dense" alternatives, i.e., arrays of the form $\{\tau_d \boldsymbol{\iota}_d : d \in \mathbb{N}\}$ for a real sequence τ_d and $\boldsymbol{\iota}_d$ denoting the vector of ones of length d, Theorem 3.5 implies that such alternatives are in $\mathscr{C}(\{\infty, \kappa_d\})$ if and only if $\sqrt{\log(d)}|\tau_d| \to \infty$; see Appendix B.1 in the Supplementary Material (Kock and Preinerstorfer (2021)) for details. Furthermore, Theorem 3.5 (together with the conditions assumed on w) immediately implies the well-known fact that $\boldsymbol{\vartheta} \in \mathscr{C}(\{\infty, \kappa_d\})$ if $\|\boldsymbol{\theta}_d\|_\infty - \sqrt{2\log(d)} \to \infty$; see Ingster and Suslina (2003, p. 92).

The discussion in the previous paragraph shows that the supremum-norm based test with asymptotic size α is consistent against the array $\{(\sqrt{3\log(d)}, 0, \ldots, 0) : d \in \mathbb{N}\}$. In contrast, it follows from Corollary 3.2 that this is not the case for any p-norm based test for $p \in (0, \infty)$. This echoes the conventional wisdom that supremum-norm based tests "are more powerful against sparse alternatives." On the other hand, Corollary 3.2 shows that for $p \in (0, \infty)$ all p-norm based tests are consistent against the array $\{\boldsymbol{\iota}_d/\sqrt{\log(d)} : d \in \mathbb{N}\}$. The above discussion implies, however, that this is not the case for the supremum-norm based test. Hence, there exists no "ranking" of the consistency sets of the p-norm based tests for $p \in (0, \infty)$ and the one of the supremum-norm based test. In particular, the monotonicity result in Theorem 3.3 does *not* extend to ∞. We summarize this observation:

Corollary 3.6. *Let* $p \in (0, \infty)$. *Then, for* $\{p, \kappa_{d,p}\}$ *and* $\{\infty, \kappa_{d,\infty}\}$ *with asymptotic sizes in* $(0, 1)$, *it holds that*

$$\mathscr{C}(\{p, \kappa_{d,p}\}) \nsubseteq \mathscr{C}(\{\infty, \kappa_{d,\infty}\}) \quad \text{and} \quad \mathscr{C}(\{\infty, \kappa_{d,\infty}\}) \nsubseteq \mathscr{C}(\{p, \kappa_{d,p}\}). \tag{20}$$

Corollary 3.6 implies in particular that supremum-norm based tests do not have the property desired in Question 1.

Although the present section has focused on the consistency set for p-norm based tests with p fixed, we conclude this section with a result concerning the consistency set of p-norm based tests with p growing with d. The following theorem relates the resulting consistency set to the one of the supremum-norm based test, and will be instrumental in the proof of Theorem 5.1 below.

Theorem 3.7. *Let the sequences* $p_d \in (0, \infty)$ *and* κ_d *be such that the sequence of tests* $\{p_d, \kappa_d\} := \mathbb{1}\{\|\cdot\|_{p_d} \geq \kappa_d\}$ *is in* $\mathbb{T}_\alpha$, $\alpha \in (0, 1)$. *Under the condition that*

$$\liminf_{d \to \infty} \frac{p_d}{\log(d)} > 2, \tag{21}$$

it holds that

$$\mathscr{C}(\{\infty, \kappa_{d,\infty}\}) \subseteq \mathscr{C}(\{p_d, \kappa_d\}) \tag{22}$$

for every sequence of critical values $\kappa_{d,\infty}$ *such that* $\{\infty, \kappa_{d,\infty}\}$ *has asymptotic size in* $(0, 1)$.

The proof of Theorem 3.7 uses the fact that $\mathscr{C}(\{\infty, \kappa_d\})$ does not depend on the concrete value of the asymptotic size, as long as this size is in $(0, 1)$, which we established in Theorem 3.5. Furthermore, it uses a "squeezing-argument" based on results in Bogachev (2006). That article provides asymptotic approximations for p_d-norm based test statistics under the null and for sequences $p_d \to \infty$ (under certain assumptions on the sequences, which necessitate the squeezing argument), cf. also Schlather (2001) and Janßen (2010) for related results. A general version of Theorem 3.7 not assuming Gaussianity is provided in Proposition C.7 and Remark C.4 in Appendix C.2 in the Supplementary Material (Kock and Preinerstorfer (2021)). Although Theorem 3.7 proves that p_d-norm based tests weakly dominate supremum-norm based tests, the discussion in Section 4.5 of Giessing and Fan (2020) reveals that tests based on a sequence of powers p_d with an asymptotic behavior as in the theorem just given do not have good power properties against dense alternatives.

4. Power enhancements and related procedures

To maximize power against sparse alternatives, it is often suggested to use a test based on the supremum-norm. At the other extreme, the typical choice maximizing power against dense alternatives is the likelihood ratio test. Therefore, one could hope that the supremum-norm based test and the likelihood ratio test together "suffice" in the sense that whenever some p-norm based test is consistent against $\boldsymbol{\vartheta}$, the supremum-norm based test or the likelihood ratio test is consistent against $\boldsymbol{\vartheta}$ as well. That is, one may conjecture that $\mathscr{C}(\{2, \kappa_{d,2}\}) \cup \mathscr{C}(\{\infty, \kappa_{d,\infty}\})$ contains

$$\bigcup_{p \in (2,\infty)} \mathscr{C}(\{p, \kappa_{d,p}\}); \tag{23}$$

all critical values being chosen so that the corresponding asymptotic sizes are in $(0, 1)$. The fact established in Theorem 3.4 that for any $q \in (2, \infty)$ all elements of the set $\mathscr{C}(\{q, \kappa_{d,q}\}) \setminus \mathscr{C}(\{2, \kappa_{d,2}\})$ are approximately sparse along a subsequence may make such a conjecture appear even more plausible, since these are the types of alternatives against which the supremum-norm based test has particularly good power properties. Based on Theorems 3.1 and 3.5 we now show, however, that the likelihood ratio test and the supremum-norm based test do *not* suffice. Specifically, we show that

$$\boldsymbol{\vartheta}^{\dagger} \in \bigcap_{p\in(2,\infty)} \mathscr{C}(\{p,\kappa_{d,p}\}) \setminus \left[\mathscr{C}(\{2,\kappa_{d,2}\}) \cup \mathscr{C}(\{\infty,\kappa_{d,\infty}\})\right], \tag{24}$$

where $\boldsymbol{\vartheta}^{\dagger} \in \boldsymbol{\Theta}$ is defined as

$$\boldsymbol{\theta}_d^{\dagger} := (\tau_d,\dots,\tau_d,0,\dots,0) \quad \text{for } \tau_d = \frac{\sqrt{2\log(d)}}{\log\log(d)}, \tag{25}$$

and where the number of non-zero entries of $\boldsymbol{\theta}_d^{\dagger}$ is $\lceil\sqrt{d}/\log(d)\rceil$ (at least for d large enough). The statement is as follows.

Theorem 4.1. *For every $p \in (2,\infty)$ and any sequences of critical values such that $\{2,\kappa_{d,2}\} \in \mathbb{T}_{\alpha_2}$, $\{\infty,\kappa_{d,\infty}\} \in \mathbb{T}_{\alpha_\infty}$, and $\{p,\kappa_{d,p}\} \in \mathbb{T}_{\alpha_p}$, the asymptotic sizes all being in $(0,1)$, it holds that*

$$\boldsymbol{\vartheta}^{\dagger} \in \mathscr{C}(\{p,\kappa_{d,p}\}) \quad \text{but} \quad \boldsymbol{\vartheta}^{\dagger} \notin \mathscr{C}(\{2,\kappa_{d,2}\}) \cup \mathscr{C}(\{\infty,\kappa_{d,\infty}\}). \tag{26}$$

Furthermore, the following convergences, stronger than the statement on the right in (26), hold

$$\mathbb{P}\left(\|\boldsymbol{\theta}_d^{\dagger} + \boldsymbol{\varepsilon}_d\|_2 \geq \kappa_{d,2}\right) \to \alpha_2 \quad \text{and} \quad \mathbb{P}\left(\|\boldsymbol{\theta}_d^{\dagger} + \boldsymbol{\varepsilon}_d\|_\infty \geq \kappa_{d,\infty}\right) \to \alpha_\infty. \tag{27}$$

Theorem 4.1 shows that there exist arrays of alternatives against which (i) *all* p-norm based tests are consistent for $p \in (2,\infty)$, and (ii) the 2- and supremum-norm based tests have asymptotic power equaling their asymptotic size.

As a corollary to this observation, we can now show that (under reasonable assumptions concerning the asymptotic sizes of the tests involved) it is impossible to obtain a test with the desired properties in Question 1 by "combining" the likelihood ratio test and the supremum-norm based test into a test φ_d such that $\mathscr{C}(\varphi_d)$ contains the union in (23). Here we say that a test φ_d is a combination of the likelihood ratio test and the supremum-norm based test if there exist sequences of critical values $\kappa_{d,2}$ and $\kappa_{d,\infty}$ such that

$$\varphi_d \leq \max\left(\{2,\kappa_{d,2}\}_d, \{\infty,\kappa_{d,\infty}\}_d\right) \quad \text{for every } d \in \mathbb{N}. \tag{28}$$

A remarkable special case of such a combination procedure was investigated in Fan, Liao and Yao (2015), where it was suggested to improve the likelihood ratio test by the supremum-norm based test in the context of their power enhancement principle. Their construction satisfies Equation (28) for critical values $\kappa_{d,2}$ such that the corresponding likelihood ratio test has asymptotic size $\alpha \in (0,1)$, and $\kappa_{d,\infty}$ such that the corresponding supremum-norm based test has asymptotic size 0. Our result concerning such procedures is as follows.

Corollary 4.2. *If $\{2,\kappa_{d,2}\} \in \mathbb{T}_{\alpha_2}$ and $\{\infty,\kappa_{d,\infty}\} \in \mathbb{T}_{\alpha_\infty}$ with $0 \leq \alpha_2 + \alpha_\infty < 1$, then $\boldsymbol{\vartheta}^{\dagger} \notin \mathscr{C}(\varphi_d)$ for every $\varphi_d \in \mathbb{T}$ that satisfies Equation (28).*

Thus, any combination of the likelihood ratio test and the supremum-norm based test is inconsistent against $\boldsymbol{\vartheta}^{\dagger}$ and hence cannot be used to answer Question 1 in the affirmative.

Remark 4.1. The array $\boldsymbol{\vartheta}^{\dagger}$ in (25) is such that $\boldsymbol{\vartheta}^{\dagger} \in \mathscr{C}(\{p, \kappa_{d,p}\})$ for all $p \in (2, \infty)$. This may give the (wrong) impression that any single p-norm based test with $p \in (2, \infty)$ can be used to detect all deviations from the null that neither the likelihood ratio test nor the supremum-norm based test are consistent against. To see that this is not the case, let $q \in [2, \infty)$ and define

$$\boldsymbol{\theta}_d^{(q)} := (\tau_d, \ldots, \tau_d, 0, \ldots, 0) \quad \text{for } \tau_d = \frac{\sqrt{2\log(d)}}{\log\log(d)},$$

but where now the number of non-zero entries of $\boldsymbol{\theta}_d^{(q)}$ is $\lceil \sqrt{d}/\log(d)^{q/2} \rceil$ (at least for d large enough). Then, a simple modification of the proof of Theorem 4.1 shows that

1. $\mathbb{P}(\|\boldsymbol{\theta}_d^{(q)} + \boldsymbol{\varepsilon}_d\|_p \geq \kappa_{d,p}) \to \alpha_p$ for every $\{p, \kappa_{d,p}\} \in \mathbb{T}_{\alpha_p}$ with $\alpha_p \in (0,1)$ and $p \in [2, q] \cup \{\infty\}$.
2. $\boldsymbol{\vartheta}^{(q)} \in \mathscr{C}(\{p, \kappa_{d,p}\})$ for every $\{p, \kappa_{d,p}\} \in \mathbb{T}_{\alpha_p}$ with $\alpha_p \in (0,1)$ and $p \in (q, \infty)$.

We observe that $\boldsymbol{\vartheta}^{(2)}$ coincides with $\boldsymbol{\vartheta}^{\dagger}$.

5. Tests that dominate all p-norm based tests

We now answer Question 1 in the affirmative by constructing sequences of tests ψ_d that are consistent against any deviation from the null that some p-norm based test is consistent against. A similar statement also holds in the non-Gaussian case, cf. Appendix C.3 in the Supplementary Material (Kock and Preinerstorfer (2021)).

The idea underlying our construction is related to the test proposed in Xu et al. (2016), who went beyond the classical power enhancement principle and suggested to combine a *fixed* number of p-norm based tests. Importantly, however, we intend to combine *all* p-norm based tests into a better test. In principle, this would require us to combine an *uncountable* amount of tests indexed by p in the non-compact set $(0, \infty)$, which seems impossible at first sight. But the monotonicity result in Theorem 3.3 can be used as a "discretization device," allowing us to get all consistency sets corresponding to powers $p \in (0, \infty)$ by suitably combining a finite, but in d *increasing*, number m_d, say, of powers $p_1, \ldots, p_{m_d}$, say, cf. Equation (31) below. Besides the monotonicity result, a crucial aspect exploited in the proof is the independence of the consistency set of p-norm based tests of their asymptotic size, cf. Theorems 3.1 and 3.5 and the ensuing discussions.

Theorem 5.1. *Let p_d be a strictly increasing and unbounded sequence in $(0, \infty)$ and let m_d be a non-decreasing and unbounded sequence in $\mathbb{N}$. Choose $\alpha \in (0,1)$ and fix an array*

$$\mathcal{A} = \left\{\alpha_{d,j} \in (0,1) : d \in \mathbb{N},\ j = 1, \ldots, m_d\right\}$$

such that, for $\mathbb{M} \subseteq \mathbb{N}$ unbounded, it holds that

$$\sum_{j=1}^{m_d} \alpha_{d,j} = \alpha \text{ for every } d \in \mathbb{N}, \quad \lim_{d\to\infty} \alpha_{d,m_d} > 0, \quad \text{and} \quad \lim_{d\to\infty} \alpha_{d,j} > 0 \text{ for every } j \in \mathbb{M}, \tag{29}$$

where the conditions implicitly impose the existence of the respective limits. For every $d \in \mathbb{N}$ and every $j = 1, \ldots, m_d$, choose $\kappa_{d,p_j} > 0$ and $c_d \in (0,1]$ such that

$$\mathbb{P}\left(\|\boldsymbol{\varepsilon}_d\|_{p_j} \geq \kappa_{d,p_j}\right) = \alpha_{d,j} \quad \text{and} \quad \mathbb{E}\left(\psi_d(\boldsymbol{\varepsilon}_d)\right) = \alpha, \tag{30}$$

where

$$\psi_d(\cdot) := \mathbb{1}\left\{\max_{j=1,\ldots,m_d} \kappa_{d,p_j}^{-1}\|\cdot\|_{p_j} \geq c_d\right\}. \tag{31}$$

Then, the following statements hold:

1. *The sequence of tests ψ_d has the property requested in Question 1, that is, it has asymptotic size α and satisfies*

$$\mathscr{C}(\{p, \overline{\kappa}_{d,p}\}) \subseteq \mathscr{C}(\psi_d), \textit{ for every } p \in (0,\infty) \textit{ and every } \{p, \overline{\kappa}_{d,p}\} \in \mathbb{T}_\alpha. \tag{32}$$

2. *Under the additional condition that*

$$\liminf_{d\to\infty} \frac{p_{m_d}}{\log(d)} > 2, \tag{33}$$

it furthermore holds that

$$\mathscr{C}(\{\infty, \kappa_d\}) \subseteq \mathscr{C}(\psi_d), \textit{ for every } \{\infty, \kappa_d\} \in \mathbb{T}_\alpha.$$

Remark 5.1. Let us note first that for every $d \in \mathbb{N}$ the acceptance region of the test ψ_d in (31) is a symmetric convex set if $p_1 \geq 1$. It then follows that ψ_d is admissible and unbiased (cf. Birnbaum (1955) and Stein (1956), and Anderson (1955)).

Second, in case (33) is satisfied, the test ψ_d does not only have the property of being consistent against every array of alternatives that any p-norm based test with $p \in (0,\infty)$ and asymptotic size α is consistent against. It moreover also dominates any supremum-norm based test with asymptotic size α. The proof of this property crucially relies on Theorem 3.7.

Third, by Theorems 3.1 and 3.5, the consistency sets of p-norm based tests with asymptotic size in $(0,1)$ do not depend on the actual value of the asymptotic size. It therefore follows that an apparently stronger (but equivalent) version of Theorem 5.1 is true, in which $\mathbb{T}_\alpha$ in Parts 1 and 2 is replaced by $\bigcup_{\tilde{\alpha}\in(0,1)} \mathbb{T}_{\tilde{\alpha}}$.

The sequence m_d regulates the number of norms the test in Equation (31) is based on, whereas the sequence p_d determines the concrete powers $p \in (0,\infty)$ used in the construction. The condition in Equation (33) requires that the maximal power used in the test (31) grows sufficiently quickly in d to guarantee that the supremum-norm based test is dominated by making use of Theorem 3.7.

Intuitively, the role of the array $\mathscr{A}$ is to regulate the sizes of the "individual" tests involved in the construction in Equation (31). Furthermore, c_d is a correction term guaranteeing that the test in that display has size *exactly* equal to α for every sample size d. The choice of $c_d = 1$ would in general only lead to a test of size *not greater* than α. Thus, working with a smaller c_d leads to higher power compared to $c_d = 1$, which would correspond to an overly conservative test. The critical values κ_{d,p_j} and the multiplier c_d as in Theorem 5.1 can be found by a simple line search. The probabilities that need to be obtained in such computations can be approximated numerically. This is computationally relatively cheap, because the number m_d of tests involved can be chosen to grow slowly in d, cf. Example 5.1 below.

Remark 5.2. Given sequences p_d and m_d as in Theorem 5.1, concrete choices of arrays $\mathscr{A}$ satisfying $\sum_{j=1}^{m_d} \alpha_{d,j} = \alpha$ for every $d \in \mathbb{N}$, and Equation (29) in Theorem 5.1 with $\mathbb{M} = \mathbb{N}$, can easily be obtained from any probability mass function $\delta_j > 0$, $j \in \mathbb{N}$, and $\gamma \in (0,1)$, via the transformation (for d large enough so that $m_d \geq 2$)

$$\alpha_{d,j} = \begin{cases} \gamma\alpha \dfrac{\delta_j}{\sum_{i=1}^{m_d-1} \delta_i} & \text{for } j = 1, \dots, m_d - 1, \\ (1-\gamma)\alpha & \text{for } j = m_d. \end{cases} \tag{34}$$

Example 5.1. A specific example of a test as in Theorem 5.1 is given next. The test we discuss commences with the likelihood ratio test in the sense that $p_1 = 2$ (this will be important in Example 5.2 below). Choose $\mathcal{A}$ as in Remark 5.2 with the geometric probability mass function $\delta_j = \delta_0(1-\delta_0)^{j-1}$ for some $\delta_0 \in (0,1)$ and $\gamma \in (0,1)$. Choose $m_d = \lceil 3\log(d) \rceil + 1$, and let $p_d = d + 1$, so that (33) is satisfied. For critical values κ_{d,p_j} as defined in (30) (but based on the concrete array, and the concrete sequences p_d and m_d just defined) and the corresponding $c_d \leq 1$, Theorem 5.1 shows that the corresponding sequence of tests satisfies the property sought for in Question 1 (and also dominates any supremum-norm based test).

The array $\mathcal{A}$ that the test ψ_d in (31) is based on also regulates the uniform "closeness" of the asymptotic power function of ψ_d to the power function of each p_j-norm based test involved in its construction. The following result quantifies this relation. How the result can be used to mimic the asymptotic power properties of a specific p-norm based test when working with a test ψ_d will be discussed subsequently.

Theorem 5.2. *In the context of Theorem 5.1, fix $j \in \mathbb{M}$ and set*

$$\lim_{d\to\infty} \alpha_{d,j} =: \underline{\alpha}_j > 0. \tag{35}$$

For any sequence of critical values κ_d such that $\{p_j, \kappa_d\} \in \mathbb{T}_\alpha$, the sequence of tests ψ_d as defined in (31) satisfies

$$\limsup_{d\to\infty} \sup_{\boldsymbol{\theta}_d \in \mathbb{R}^d} \left[\mathbb{P}\left(\|\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d\|_{p_j} \geq \kappa_d \right) - \mathbb{E}(\psi_d(\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d)) \right] \leq \frac{\Phi^{-1}(1-\underline{\alpha}_j) - \Phi^{-1}(1-\alpha)}{\sqrt{2\pi}}, \tag{36}$$

and

$$\lim_{d\to\infty} \sup_{\boldsymbol{\theta}_d \in \mathbb{R}^d} \left[\mathbb{E}(\psi_d(\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d)) - \mathbb{P}\left(\|\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d\|_{p_j} \geq \kappa_d \right) \right] = 1 - \alpha. \tag{37}$$

Recall from Equation (32) that the asymptotic power of ψ_d is 1 whenever that of some p-norm based test is. The inequality in (36) sheds further light on the asymptotic power function of ψ_d by establishing that it is *nowhere* much below that of $\{p_j, \kappa_d\}$ for an index $j \in \mathbb{M}$ such that $\underline{\alpha}_j \approx \alpha$, as the upper bound in that inequality is then approximately 0. Thus, there are no arrays of alternatives for which much is lost by using ψ_d instead of such a $\{p_j, \kappa_d\}$. What is more, Equation (37) (and its proof) shows that much can be gained by using ψ_d, as arrays of alternatives exist against which ψ_d is consistent, but against which $\{p_j, \kappa_d\}$ has asymptotic power equal to its size.

In the following example we illustrate how this reasoning can be incorporated in the construction of ψ_d.

Example 5.2. In case one has reasons to favor a specific p-norm based test, e.g., the likelihood ratio test corresponding to $p = 2$, but does not want to abandon the idea of using a test ψ_d that dominates all p-norm based tests, one can decide on the following compromise, which is possible as a consequence of Theorems 5.1 and 5.2: choose the components in the construction of ψ_d in such a way that ψ_d dominates all p-norm based tests *and* such that the power function of the obtained test ψ_d is *everywhere* at most slightly smaller than that of the preferred p-norm based test.

To see how this can be achieved, we focus on the case where $p = 2$. Let $\{2, \kappa_d\}$ be a sequence of likelihood ratio tests with asymptotic size $\alpha \in (0,1)$. To obtain a test ψ_d as in Theorem 5.1 and whose power is (asymptotically) nowhere much less than that of $\{2, \kappa_d\}$, we can reconsider the test ψ_d constructed in Example 5.1. Note that $p_1 = 2$. Furthermore, $\underline{\alpha}_1 = \alpha\gamma\delta_0$, cf. Equation (35). It hence follows from Theorem 5.2 (with $j = 1$) that

$$\limsup_{d\to\infty} \sup_{\boldsymbol{\theta}_d \in \mathbb{R}^d} \left[\mathbb{P}\left(\|\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d\|_2 \geq \kappa_d\right) - \mathbb{E}(\psi_d(\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d))\right] \leq \frac{\Phi^{-1}(1-\alpha\gamma\delta_0) - \Phi^{-1}(1-\alpha)}{\sqrt{2\pi}}.$$

The upper bound can be made arbitrarily close to 0 by choosing δ_0 and γ close to 1. At the same time, and in contrast to the likelihood ratio test $\{2, \kappa_d\}$, ψ_d has the favorable property of being consistent against every alternative that some p-norm based test is consistent against.

6. Minimax adaptive testing

The main focus of the present article is to study the consistency behavior of p-norm based tests for the *unrestricted* testing problem (2), and to answer questions like "how do these tests compare" and "can they be dominated" in terms of their consistency properties.

Classical results on p-norm based tests in the literature on minimax-optimal testing, on the other hand, start with a specific set of alternatives, i.e., complements of p-norm balls centered at the origin. They then characterize the separation from the null necessary so that uniform consistency is possible, and show that a p-norm based test can be constructed that is minimax rate optimal against such alternatives. Such results provide a justification for using a p-norm based test if one cares about power against alternatives in the complement of a p-norm ball as just described, but they do not answer the questions that we have focused on in previous sections.

In the present section, we want to illustrate that tools similar to the ones used in previous sections can be used to establish adaptivity results in the minimax framework. Essentially, we show that a p-norm based test is not only minimax optimal against complements of centered p-norm balls, but is adaptively minimax optimal over all complements of centered q-norm balls with $q \leq p$ (assuming that the radii of the balls admit minimax consistent tests), thus extending a result in Ingster and Suslina (2003) for the case where $p \leq 2$. This corresponds to the monotonicity result in Theorem 3.3. We also prove an adaptivity result over the whole range of all $p \in (0, \infty)$ which parallels Theorem 5.1.

Let us first revisit classical results and introduce some notation. Given $p \in (0,\infty)$ and a radius $r \in (0,\infty)$, let $\mathbb{V}_p^d(r) := \{\boldsymbol{\theta}_d \in \mathbb{R}^d : \|\boldsymbol{\theta}_d\|_p \geq r\}$. For a sequence of such radii $r_{p,d}$, $d \in \mathbb{N}$, we now consider (for every $d \in \mathbb{N}$) the testing problem

$$H_{0,d} : \boldsymbol{\theta}_d = \mathbf{0}_d \quad \text{vs.} \quad H_{1,d} : \boldsymbol{\theta}_d \in \mathbb{V}_p^d(r_{p,d}), \tag{38}$$

and denote the minimal sum of Type 1 and Type 2 errors for this testing problem by

$$\gamma(r_{p,d}, p) := \inf_{\varphi_d} \left[\mathbb{E}\varphi_d(\boldsymbol{\varepsilon}_d) + 1 - \inf_{\boldsymbol{\theta}_d \in \mathbb{V}_p^d(r_{p,d})} \mathbb{E}\varphi_d(\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d)\right]; \tag{39}$$

here the outer infimum is taken over all Borel measurable functions from $\mathbb{R}^d$ to $[0,1]$.

Compared to the testing problem (2) studied in previous sections, the alternatives considered in (38) are now separated from the null, the type of separation depending on p. The main questions concerning the sequence of testing problems (38) (which have long been answered) are: (i) for which sequences of radii does $\gamma(r_{p,d}, p) \to 0$; and (ii) for which tests this is achieved.

Following Ingster and Suslina (2003), a sequence of radii $r^*_{p,d}$, $d \in \mathbb{N}$, is called a sequence of asymptotic *minimax rates/critical radii* for the sequence of testing problems (38) if the following holds for any sequence of radii $r_{p,d}$ as above and as $d \to \infty$:

$$\gamma(r_{p,d}, p) \to 0 \quad \text{if and only if} \quad r_{p,d}/r^*_{p,d} \to \infty$$

$$\gamma(r_{p,d}, p) \to 1 \quad \text{if and only if} \quad r_{p,d}/r^*_{p,d} \to 0.$$

Proposition 3.9 (and its proof) in Ingster and Suslina (2003) settles questions (i) and (ii) above as follows: (i) The following sequences constitute sequences of critical radii

$$r^*_{p,d} = \begin{cases} d^{\frac{4-p}{4p}} & \text{if } p \in (0,2] \\ d^{\frac{1}{2p}} & \text{if } p \in (2,\infty). \end{cases} \tag{40}$$

(ii.a) For every $p \in [2,\infty)$ and for every sequence of radii $r_{p,d}$ such that $r_{p,d}/r^*_{p,d} \to \infty$, there exists a sequence of critical values $\kappa_{p,r_{p,d},d}$, such that

$$\mathbb{P}\left(\|\boldsymbol{\varepsilon}_d\|_p \geq \kappa_{p,r_{p,d},d}\right) + 1 - \inf_{\boldsymbol{\theta}_d \in \mathbb{V}^d_p(r_{p,d})} \mathbb{P}\left(\|\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d\|_p \geq \kappa_{p,r_{p,d},d}\right) \to 0.$$

That is, there exists a p-norm based test (the critical values depending on the sequence of radii $r_{p,d}$) that is *minimax rate consistent* in the sequence of testing problems (38).[4] (ii.b) For every $p \in (0,2)$ and for every sequence of radii $r_{p,d}$ such that $r_{p,d}/r^*_{p,d} \to \infty$, there exists a 2-norm based test that is minimax rate consistent in the sequence of testing problems (38).

Inspection of the proof of Proposition 3.9 and using Corollary 3.4 in Ingster and Suslina (2003) shows that actually more can be said in case (ii.b): Given a set of radial sequences

$$\{r_{q,d} \in (0,\infty) : q \in (0,2], d \in \mathbb{N}\} \quad \text{such that} \quad \inf_{q \in (0,2]} r_{q,d}/r^*_{q,d} \to \infty,$$

there exists a 2-norm based test that is minimax rate consistent for the sequence of testing problems

$$H_{0,d} : \boldsymbol{\theta}_d = \mathbf{0}_d \quad \text{vs.} \quad H_{1,d} : \boldsymbol{\theta}_d \in \bigcup_{q \in (0,2]} \mathbb{V}^d_q(r_{q,d}); \tag{41}$$

that is, the 2-norm based test *adapts* to $q \in (0,2]$. The following theorem now shows that such a result actually extends from the 2-norm based test to every $p \in [2,\infty)$, which parallels the monotonicity phenomenon in Theorem 3.3. Throughout the remainder of this section, we shall again use the notation $\sigma^2_p := \mathbb{V}ar(|\varepsilon_1|^p)$ and $\mu_p := \mathbb{E}(|\varepsilon_1|^p)$.

---

[4] In light of the minimax rate being $r^*_{p,d} = d^{\frac{4-p}{4p}}$ for $p \in (0,2)$ it is tempting to conjecture that (12) could be replaced by $\boldsymbol{\vartheta} \in \mathscr{C}(\{p,\kappa_d\})$ being equivalent to $\frac{\|\boldsymbol{\theta}_d\|_2^2}{\sqrt{d}} \wedge \frac{\|\boldsymbol{\theta}_d\|_p^p}{d^{\frac{4-p}{4}}} \to \infty$. This, however, is not the case as can be seen by considering $\boldsymbol{\theta}_d = (d^{\frac{1}{2p}}\log(d), 0, \ldots, 0)$.

Theorem 6.1. *Fix $p \in (2,\infty)$ and suppose a set of radial sequences*

$$\{r_{q,d} \in (0,\infty) : q \in (0,p],\ d \in \mathbb{N}\} \quad \textit{satisfies} \quad r_d := \inf_{q \in (0,p]} r_{q,d}/r^*_{q,d} \to \infty.$$

Then, the sequence of tests $\{p, \kappa_d\}$ with $\kappa_d := \left[r_d\sqrt{d\sigma_p^2} + d\mu_p\right]^{1/p}$ is minimax rate consistent in the sequence of testing problems

$$H_{0,d} : \boldsymbol{\theta}_d = \mathbf{0}_d \quad \textit{vs.} \quad H_{1,d} : \boldsymbol{\theta}_d \in \mathbb{V}_{p,d} := \bigcup_{q \in (0,p]} \mathbb{V}^d_q(r_{q,d}); \tag{42}$$

that is,

$$\mathbb{P}\left(\|\boldsymbol{\varepsilon}_d\|_p \geq \kappa_d\right) + 1 - \inf_{\boldsymbol{\theta}_d \in \mathbb{V}_{p,d}} \mathbb{P}\left(\|\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d\|_p \geq \kappa_d\right) \to 0. \tag{43}$$

Given the adaptivity result just obtained, one may ask whether one can construct a *single* test that is minimax rate consistent in any testing problem of the type (42). That is, does there exist a *single* test that is minimax rate consistent in the sequence of testing problems (42) for *every* $p \in (0,\infty)$ simultaneously? As we shall establish next, this question can be answered affirmatively using a construction related to the one in Theorem 5.1.

Theorem 6.2. *Suppose a set of radial sequences*

$$\{r_{q,d} \in (0,\infty) : q \in (0,\infty),\ d \in \mathbb{N}\} \quad \textit{satisfies} \quad r_d := \inf_{q \in (0,\infty)} r_{q,d}/r^*_{q,d} \to \infty.$$

Let p_d be a non-decreasing and diverging sequence of natural numbers satisfying

$$p_d\Phi(-r_d) \to 0 \quad \textit{and} \quad \frac{p_d}{\sqrt{d}}(3/2)^{3p_d/2} \to 0. \tag{44}$$

Then, setting $\kappa_{d,j} := \left[r_d\sqrt{d\sigma_j^2} + d\mu_j\right]^{1/j}$, the sequence of tests

$$\psi^*_d(\cdot) := \mathbb{1}\left\{\max_{j=1,\dots,p_d} \kappa^{-1}_{d,j}\|\cdot\|_j \geq 1\right\} \tag{45}$$

is minimax rate consistent in the sequence of testing problems (42) for every $p \in (0,\infty)$; that is,

$$\mathbb{E}\psi^*_d(\boldsymbol{\varepsilon}_d) + 1 - \inf_{\boldsymbol{\theta}_d \in \mathbb{V}_{p,d}} \mathbb{E}\psi^*_d(\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d) \to 0 \qquad \textit{for every } p \in (0,\infty). \tag{46}$$

For concreteness, we have decided to base the construction in (45) on all powers from 1 to p_d. Inspection of the proof shows that this is not crucial, and that one can also achieve the same minimax rate consistency property as in Theorem 6.2 by maximizing over an expanding subset of powers.

7. Numerical results

To investigate the non-asymptotic properties of the tests under consideration, we provide a numerical comparison of the power of p-norm based tests (we consider $p = 1, 2, 3, 4$ and $p = \infty$) and a test ψ_d, say, corresponding to the construction in Theorem 5.1. We set $\alpha = 0.05$ and consider the dimensions $d = 50.000$ and $d = 250.000$. The specific version of ψ_d used in the computations is the following:

1. We employed $m_d = \lceil \log(\log(d^6)) \rceil$, which equals 5 for both dimensions considered.
2. We used $p_d = \exp(d-1)+1$, i.e., the test was based on $p_1 = 2$, $p_2 = e+1$, $p_3 = e^2+1$, $p_4 = e^3+1$, and $p_5 = p_{m_d} = e^4+1$.
3. To generate $\mathscr{A}$, we used the approach in Remark 5.2 with $\gamma = 1/2$ and δ_j $(j = 1, \dots, m_d - 1)$ the probability mass function from a geometric distribution with success parameter $\delta_0 = 1/2$. For both dimensions considered, this results in roughly $\alpha_{d,1} = 0.013$, $\alpha_{d,2} = 0.007$, $\alpha_{d,3} = 0.003$, $\alpha_{d,4} = 0.002$ and $\alpha_{d,5} = \alpha_{d,m_d} = 0.025$.

Note that with this choice Equation (33) holds and that m_d, i.e., the number of exponents used in the construction, grows slowly with d which is numerically favorable. As a consequence of Theorem 5.2, and since $p_1 = 2$, we obtain that the asymptotic power of the Euclidean norm based test can nowhere exceed the asymptotic power of ψ_d by more than $(\Phi^{-1}(0.9875) - \Phi^{-1}(0.95))/\sqrt{2\pi} \approx 0.24$ (noting that $1 - \gamma\alpha\delta_0 = 0.9875$, cf. Example 5.2). The critical values for the tests considered were obtained through Monte Carlo (with 50.000 replications throughout).

We compare the power functions (determined via Monte Carlo using 1.000 replications) of each of the above-mentioned tests against three types of alternatives: (i) dense vectors $\boldsymbol{\theta}_d$, i.e., vectors of the form $a \times (1, \dots, 1)$; (ii) semi-sparse vectors $\boldsymbol{\theta}_d$, i.e., vectors of the form $a \times \boldsymbol{\theta}_d^{\dagger}$ for $\boldsymbol{\theta}_d^{\dagger}$ as defined in Equation (25); and (iii) sparse vectors $\boldsymbol{\theta}_d$, i.e., vectors of the form $a \times (1, 0, \dots, 0)$. The power functions are provided in Figure 3 and are plotted against a. The results show that the choice of the exponent in a p-norm based test has a strong effect on the type of signal one has high power against. Whereas the supremum-norm based test performs well for sparse and semi-sparse signals, lower exponents perform better for dense signals, and vice versa. The combination procedure ψ_d strikes a balance between this extreme difference in performance. It performs best in the semi-sparse setup, while it is very competitive with the best performing tests in the other setups, particularly so for values of a where power is high, echoing our theoretical results.

8. Conclusion

Combining two tests to enhance power is the underlying paradigm of the power enhancement principle recently put forward by Fan, Liao and Yao (2015). There, it was suggested to combine tests based on the Euclidean and supremum-norms. In the present article, we have characterized the consistency sets of p-norm based tests for all $p \in (0, \infty]$. Our characterizations allowed us to reveal an unexpected monotonicity relation of the consistency sets, and, somewhat surprisingly, to asymptotically dominate *all* these tests in terms of set inclusion of consistency sets, but also in a minimax sense. This was achieved by suitably combining a finite number of tests that grows with the dimension of the testing problem.

Even though the Gaussian sequence model is a prototypical framework for high-dimensional inference and results in that setup carry over to many other settings at least on a conceptual level, suitable generalizations of our results continue to hold also in non-Gaussian settings, as we show in detail in Appendix C in the Supplementary Material (Kock and Preinerstorfer (2021)).

Acknowledgements

We would like to thank Davy Paindaveine, Michael Wolf, seminar participants in Bristol/Warwick, Durham, Ohio, Turku, Paris, Rome, Vienna, Wisconsin, York, and a referee for helpful comments

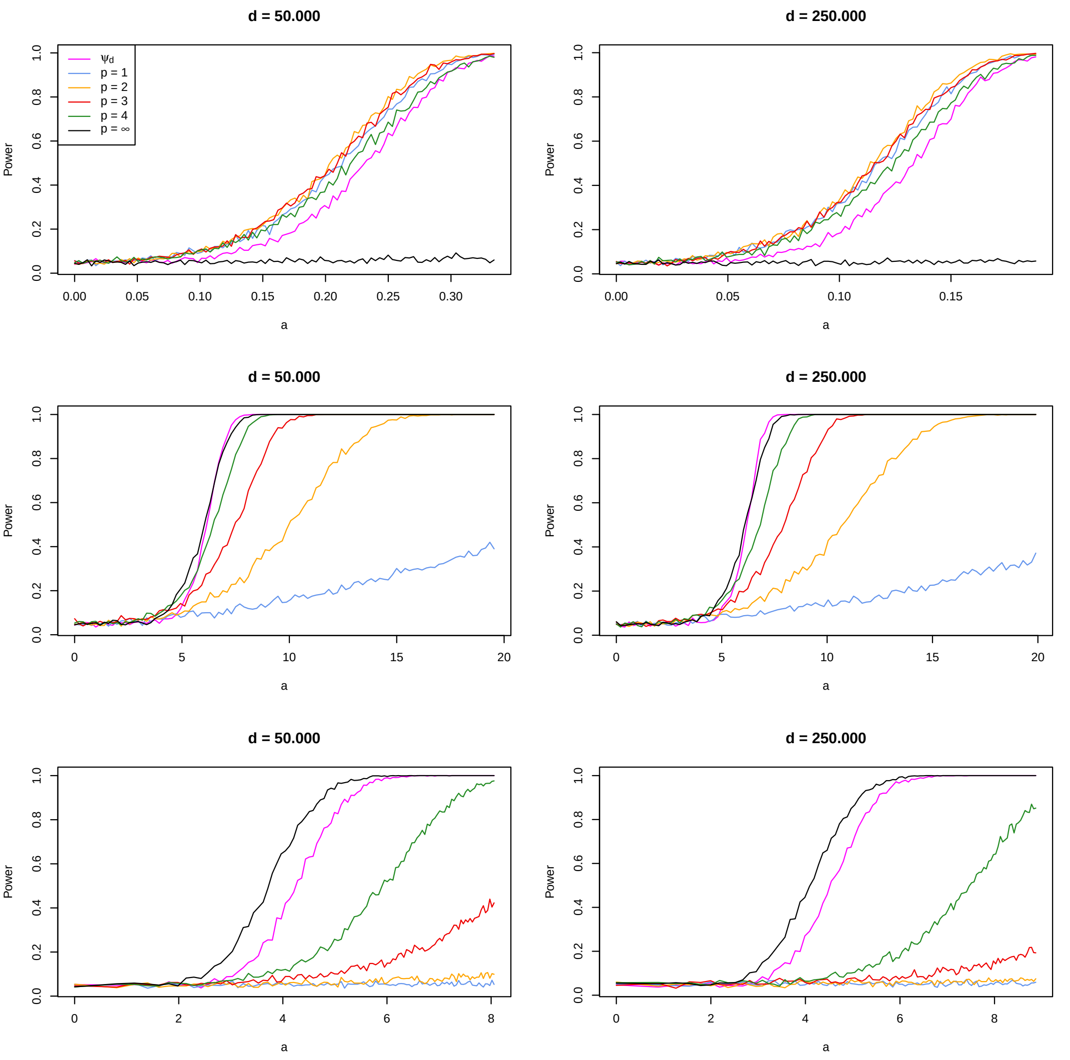

Figure 3: Power against dense alternatives (first row), semi-sparse alternatives (second row), and sparse alternatives (last row).

and discussions. David Preinerstorfer gratefully acknowledges support from the Program of Concerted Research Actions (ARC) of the Université libre de Bruxelles.

Supplementary material

The Supplementary Material (Kock and Preinerstorfer (2021)) contains Appendices B and C.

Appendix A: Proofs, Part I

We here collect the proofs for the results in the main body of this article, excluding the proofs for the results in Section 6 and the proof of a claim in Section 3.2. These results are established in Appendix B in the Supplementary Material (Kock and Preinerstorfer (2021)). In the following, we shall freely apply results from Appendix C ("Supplementary general results *not* imposing Gaussianity") without mentioning each time that Appendix C can be found in the Supplementary Material as well. Throughout Appendix A, all assumptions imposed in the main body of the paper concerning the distribution of the random variables ε_i are maintained; that is, the ε_i are i.i.d. standard normal.

A.1. Proof of Theorem 2.1

We establish a slightly stronger result than Theorem 2.1, showing that the sequence of tests $\overline{\psi}_d$ in Theorem 2.1 can actually be constructed such that the power of the sequence of tests ψ_d is improved against a 1-sparse vector of alternatives. This is of additional interest, as it implies that the result pertains even in situations where the testing problem is reduced to 1-sparse parameter spaces, i.e., when instead of considering the parameter space $\boldsymbol{\Theta} = \bigtimes_{d=1}^{\infty} \mathbb{R}^d$ one only considers all arrays $\boldsymbol{\vartheta}$ in

$$\boldsymbol{\Sigma}_1 := \bigtimes_{d=1}^{\infty} \bigcup_{i=1}^{d} \left\{\theta \boldsymbol{e}_i(d) : \theta \in \mathbb{R}\right\} \subseteq \boldsymbol{\Theta},$$

where $\boldsymbol{e}_i(d)$ denotes the i-th element of the canonical basis of $\mathbb{R}^d$. Note that for every fixed $d \in \mathbb{N}$, every element of $\bigcup_{i=1}^{d} \left\{\theta \boldsymbol{e}_i(d) : \theta \in \mathbb{R}\right\}$ has at most one non-zero coordinate, i.e., is 1-sparse.

Theorem A.1. *Let $\alpha \in (0,1)$. For every $\psi_d \in \mathbb{T}_\alpha$ there exists a $\overline{\psi}_d \in \mathbb{T}_\alpha$, such that:*

1. *$\overline{\psi}_d \geq \psi_d$ holds for every $d \in \mathbb{N}$, guaranteeing that $\overline{\psi}_d$ has uniformly non-inferior asymptotic power compared to ψ_d;*
2. *$\overline{\psi}_d$ is consistent against an array $\boldsymbol{\vartheta} \in \boldsymbol{\Sigma}_1$ against which ψ_d has asymptotic power at most α.*

In particular it holds that $\mathscr{C}(\psi_d) \subsetneqq \mathscr{C}(\overline{\psi}_d)$.

The proof is a special case of the example discussed in Section 1.1 of Kock and Preinerstorfer (2019). The simplification is due to no sufficiency argument being necessary in the present framework. We provide a complete argument for the convenience of the reader.

Proof. The first part of the proof is based on an argument in Section 3.4.2 of Ingster and Suslina (2003): For every $d \in \mathbb{N}$, set $\mathbb{Q}_{d,0} := \mathbb{N}(\boldsymbol{0}_d, \boldsymbol{I}_d)$ (the distribution of $\boldsymbol{y}_d$ under the null) and define the mixture $\mathbb{Q}_d = d^{-1} \sum_{i=1}^{d} \mathbb{Q}_{d,i}$, where $\mathbb{Q}_{d,i} := \mathbb{N}(a_d \boldsymbol{e}_i(d), \boldsymbol{I}_d)$ (the distribution of $\boldsymbol{y}_d$ under the 1-sparse alternative $\boldsymbol{\theta}_d = a_d \boldsymbol{e}_i(d)$) for $a_d := \sqrt{\log(d)/2}$. The likelihood-ratio statistic of $\mathbb{Q}_d$ w.r.t. $\mathbb{Q}_{d,0}$ is given by $L_d(z_1, \ldots, z_d) = d^{-1} \sum_{i=1}^{d} e^{a_d z_i - a_d^2/2}$. Denote the expectation operators w.r.t. $\mathbb{Q}_d$ and $\mathbb{Q}_{d,i}$ by $\mathbb{E}_d^Q$ and $\mathbb{E}_{d,i}^Q$, respectively. It holds that

$$\left|\mathbb{E}_{d,0}^Q(\psi_d) - d^{-1} \sum_{i=1}^{d} \mathbb{E}_{d,i}^Q(\psi_d)\right|^2 = \left|\mathbb{E}_{d,0}^Q(\psi_d(1 - L_d))\right|^2 \leq \mathbb{E}_{d,0}^Q\left((1 - L_d)^2\right) = \mathbb{E}_{d,0}^Q\left(L_d^2\right) - 1, \tag{47}$$

where we used $d^{-1}\sum_{i=1}^{d} \mathbb{E}^{Q}_{d,i}(\psi_d) = \mathbb{E}^{Q}_{d}(\psi_d) = \mathbb{E}^{Q}_{d,0}(\psi_d L_d)$, Jensen's inequality and $\mathbb{E}^{Q}_{d,0}(L_d) = 1$. From the moment-generating-function of a normal distribution we obtain

$$\mathbb{E}^{Q}_{d,0}\left(L_d^2\right) - 1 = d^{-2} \sum_{i=1}^{d} \sum_{j=1}^{d} \mathbb{E}^{Q}_{d,0}\left(e^{a_d(z_i+z_j)-a_d^2}\right) - 1 \leq d^{-1/2} \to 0. \tag{48}$$

Given this argument, we now see that the sequence to the left in (47) converges to 0, and thus the existence of an array $\boldsymbol{\vartheta} = \{a_d \boldsymbol{e}_{i(d)}(d) : d \in \mathbb{N}\} \in \boldsymbol{\Sigma}_1$ follows, against which ψ_d has asymptotic power at most α. Next, define (for every $d \in \mathbb{N}$) the test $\nu_d = \mathbb{1}\{y \in \mathbb{R}^d : |\boldsymbol{e}'_{i(d)} y| \geq \sqrt{a_d}\}$. Obviously, its asymptotic size equals 0, and it is consistent against $\boldsymbol{\vartheta}$. Hence $\overline{\psi}_d := \min(\psi_d + \nu_d, 1)$ has the properties required in the theorem. ☐

Remark A.1. Theorem A.1 is an "existence result" in the sense that it shows that there exists a sequence of tests $\overline{\psi}_d$ with certain properties. Inspection of the proof just given, however, shows how such a sequence of tests $\overline{\psi}_d$ can actually be obtained by "enhancing" the test ψ_d with a test ν_d as defined in course of the proof. Note that the sequence $i(d)$ needed in such a construction can in principle be found numerically by choosing $i(d)$ for every $d \in \mathbb{N}$ as an index $i \in \{1, \ldots, d\}$ such that the test ψ_d has minimal power against $\mathbb{N}(a_d \boldsymbol{e}_i(d), \boldsymbol{I}_d)$ (note that the d values of the power function needed can be obtained by Monte-Carlo simulations). The same remark applies a fortiori to Theorem 2.1.

A.2. p-norm based tests with $p \in (0, \infty)$

A.2.1. Proof of Theorem 3.1

By Remark C.1, Assumption C.1 holds for $F = \Phi$, so that the theorem is an immediate consequence of Theorem C.4 (since $\mathbb{E}(|\varepsilon|^q) < \infty$ then holds for every $q \in (0, \infty)$).

A.2.2. Proof of Corollary 3.2

To show (11), recalling the definition of g_p from (9), we just combine Theorem 3.1 with $\frac{1}{2}(|\cdot|^p + |\cdot|^2) \leq g_p(\cdot) \leq |\cdot|^p + |\cdot|^2$ for $p \in [2, \infty)$. Similarly, to show (12), we use Theorem 3.1 together with $g_p(\cdot) \leq |\cdot|^p$ and $g_p(\cdot) \leq |\cdot|^2$ for $p \in (0, 2)$. To verify the remaining statement, let $\boldsymbol{\vartheta} \in \boldsymbol{\Theta}$ be defined via

$$\theta_{i,d} = \begin{cases} d^{-1/4} & \text{for } i = 1, \ldots, d-1, \\ d^{1/(2p)} & \text{for } i = d. \end{cases}$$

Then $\|\boldsymbol{\theta}_d\|_2^2/\sqrt{d} \geq d^{\frac{1}{p}-\frac{1}{2}} \to \infty$ and $\|\boldsymbol{\theta}_d\|_p^p/\sqrt{d} \geq \frac{d-1}{\sqrt{d}} d^{-p/4} \to \infty$. However,

$$d^{-1/2} \sum_{i=1}^{d} g_p(\theta_{i,d}) = \frac{d-1}{d} + 1 \leq 2,$$

and we conclude $\boldsymbol{\vartheta} \notin \mathscr{C}(\{p, \kappa_d\})$ with Theorem 3.1.

A.2.3. Proof of Theorem 3.3

Just observe that $0 < p < q < \infty$ implies $g_p \leq g_q$, which by Theorem 3.1 delivers $\mathscr{C}(\{p, \kappa_{d,p}\}) \subseteq \mathscr{C}(\{q, \kappa_{d,q}\})$. That the inclusion is strict follows from

$$\boldsymbol{\vartheta} := \{(d^{1/(2p)}, 0, \ldots, 0) : d \in \mathbb{N}\} \in \mathscr{C}(\{q, \kappa_{d,q}\}) \setminus \mathscr{C}(\{p, \kappa_{d,p}\}),$$

which is easily seen using the equivalence from Theorem 3.1.

A.2.4. *Proof of Theorem 3.4*

Fix $p < q$, $\{p, \kappa_{d,p}\}$ and $\{q, \kappa_{d,q}\}$ as in the statement of this theorem. Let $\boldsymbol{\vartheta} \in \mathscr{C}(\{q, \kappa_{d,q}\}) \setminus \mathscr{C}(\{p, \kappa_{d,p}\})$, which is possible by Theorem 3.3. From Theorem 3.1 and Remark 3.1 we conclude that for every $\delta \in (0, \infty)$ it holds that

$$G_{d,p}(\delta) := \frac{\sum_{i=1}^{d} \theta_{i,d}^2 \mathbb{1}\{|\theta_{i,d}| \leq \delta\} + \sum_{i=1}^{d} |\theta_{i,d}|^p \mathbb{1}\{|\theta_{i,d}| > \delta\}}{\sqrt{d}} = \frac{\sum_{i=1}^{d} g_p^{(\delta)}(\theta_{i,d})}{\sqrt{d}} \not\to \infty. \tag{49}$$

By (49), for every $\delta \in (0, \infty)$, we can choose a subsequence d'_δ of d (we highlight the dependence of the subsequence on δ in the proof to avoid confusion), such that $\sup_{d'_\delta} G_{d'_\delta,p}(\delta) \leq D(\delta) < \infty$ for some $D(\delta)$. Although d'_δ and $D(\delta)$ depend on δ, in general, if $p \geq 2$ these dependencies can be avoided, because $\sup_{d'_1} G_{d'_1,p}(\delta) \leq \sup_{d'_1} G_{d'_1,p}(1) =: D(1) < \infty$ then holds for all $\delta \in (0, \infty)$. To see the latter, just note that in case $p \geq 2$ the function $\delta \mapsto G_{d,p}(\delta)$, for $\delta \in (0, \infty)$, has a maximum at $\delta = 1$ for every $d \in \mathbb{N}$. In what follows we set $d'_\delta = d'_1$ and $D(\delta) = D(1)$ in case $p \geq 2$.

Irrespective of the value of $p \in (0, \infty)$, we obtain for every $\delta \in (0, \infty)$ that

$$\delta^p \times \sup_{d'_\delta} \frac{\sum_{i=1}^{d'_\delta} \mathbb{1}\{|\theta_{i,d'_\delta}| > \delta\}}{\sqrt{d'_\delta}} \leq \sup_{d'_\delta} \frac{\sum_{i=1}^{d'_\delta} |\theta_{i,d'_\delta}|^p \mathbb{1}\{|\theta_{i,d'_\delta}| > \delta\}}{\sqrt{d'_\delta}} \leq \sup_{d'_\delta} G_{d'_\delta,p}(\delta) \leq D(\delta). \tag{50}$$

In case $p \geq 2$, we can furthermore take the supremum over $\delta \in (0, \infty)$ (since the subsequence chosen does not depend on δ) to get

$$\sup_{\delta \in (0,\infty)} \delta^p \times \sup_{d'_1} \frac{\sum_{i=1}^{d'_1} \mathbb{1}\{|\theta_{i,d'_1}| > \delta\}}{\sqrt{d'_1}} \leq D(1). \tag{51}$$

Thus, the first statements in Equation (16) and (17), respectively, hold for the subsequences d'_δ and d'_1. It remains to verify that for every $\delta \in (0, \infty)$ we have that $\max_{i=1,\ldots,d'_\delta} |\theta_{i,d'_\delta}| \to \infty$ (note that $d'_\delta = d'_1$ in case $p \geq 2$ by construction). Fix $\delta \in (0, \infty)$. Theorem 3.1 and Remark 3.1 show that $\boldsymbol{\vartheta} \in \mathscr{C}(\{q, \kappa_{d,q}\})$ is equivalent to

$$\frac{\sum_{i=1}^{d} \theta_{i,d}^2 \mathbb{1}\{|\theta_{i,d}| \leq \delta\} + \sum_{i=1}^{d} |\theta_{i,d}|^q \mathbb{1}\{|\theta_{i,d}| > \delta\}}{\sqrt{d}} = \frac{\sum_{i=1}^{d} g_q^{(\delta)}(\theta_{i,d})}{\sqrt{d}} \to \infty \tag{52}$$

The last inequality in (50) delivers $\sum_{i=1}^{d'_\delta} \theta_{i,d'_\delta}^2 \mathbb{1}\{|\theta_{i,d'_\delta}| \leq \delta\} / \sqrt{d'_\delta} \leq D(\delta)$ for every d'_δ. Thus, (52) implies

$$\frac{\sum_{i=1}^{d'_\delta} |\theta_{i,d'_\delta}|^q \mathbb{1}\{|\theta_{i,d'_\delta}| > \delta\}}{\sqrt{d'_\delta}} \to \infty.$$

But for every d'_δ

$$\frac{\sum_{i=1}^{d'_\delta} |\theta_{i,d'_\delta}|^q \mathbb{1}\{|\theta_{i,d'_\delta}| > \delta\}}{\sqrt{d'_\delta}} \leq \max_{i=1,\ldots,d'_\delta} |\theta_{i,d'_\delta}|^q \frac{\sum_{i=1}^{d'_\delta} \mathbb{1}\{|\theta_{i,d'_\delta}| > \delta\}}{\sqrt{d'_\delta}},$$

which, by (50), is upper bounded by $\max_{i=1,\ldots,d'_\delta} |\theta_{i,d'_\delta}|^q D(\delta)\delta^{-p}$, which hence diverges to ∞.

A.3. Supremum-norm based tests

In some proofs in this section we will use the following classic inequality (e.g., Feller (1968), Section 7.1) for the standard normal cdf Φ:

$$\left(x^{-1}-x^{-3}\right)\frac{e^{-x^2/2}}{\sqrt{2\pi}} \leq 1-\Phi(x) \leq x^{-1}\frac{e^{-x^2/2}}{\sqrt{2\pi}} \quad \text{for every } x>0. \tag{53}$$

We also need the following observation concerning κ_d sequences leading to $\{\infty, \kappa_d\} \in \mathbb{T}_\alpha$ for some $\alpha \in (0,1)$ (which provides a more common way of writing critical values for the supremum-based test in the normal case as compared to the general case treated in Lemma C.8 in Appendix C). To obtain the result, we recall the classical limit theorem (e.g., Theorem 1.5.3 in Leadbetter, Lindgren and Rootzén (1983))

$$\mathbb{P}\left(a_d\left[\max_{1\leq i\leq d}\varepsilon_i - b_d\right]\leq x\right)\to\Lambda(x) \quad \text{for every } x\in\mathbb{R}, \tag{54}$$

where $\Lambda(x)=\exp\left(-\exp(-x)\right)$, $x\in\mathbb{R}$, denotes Gumbel's double exponential cdf, and where $a_d=\sqrt{2\log(d)}$ and $b_d=\sqrt{2\log(d)}-\frac{\log\log(d)+\log(4\pi)}{2\sqrt{2\log(d)}}$. For $x\in\mathbb{R}$ set $u_d=u_d(x):=x/a_d+b_d$, so that

$$\mathbb{P}\left(\max_{1\leq i\leq d}|\varepsilon_i|\leq u_d\right)=\mathbb{P}\left(\max_{1\leq i\leq d}\varepsilon_i\leq u_d,\ \min_{1\leq i\leq d}\varepsilon_i\geq -u_d\right)\to\exp(-2\exp(-x)),$$

where we used (54) and the asymptotic independence of the minimum and maximum (e.g., Theorem 1.8.2 in Leadbetter, Lindgren and Rootzén (1983)). That is,

$$\mathbb{P}\left(a_d\left[\max_{1\leq i\leq d}|\varepsilon_i| - b_d\right]\leq x\right)\to\exp(-2\exp(-x)) \quad \text{for every } x\in\mathbb{R}. \tag{55}$$

The following is an immediate consequence of (55).

Lemma A.2. *It holds that* $\{\infty,\kappa_d\}\in\mathbb{T}_\alpha$, $\alpha\in(0,1)$, *if and only if*

$$\kappa_d=\sqrt{2\log(d)}-\frac{\log\log(d)+\log(4\pi)}{2\sqrt{2\log(d)}}-\frac{\log\left(-\log(1-\alpha)/2\right)}{\sqrt{2\log(d)}}+o\left(\frac{1}{\sqrt{\log(d)}}\right). \tag{56}$$

We can now proceed to proving the characterization of the consistency set of supremum-norm based tests. Throughout the proof we interpret sums over empty index sets as 0.

A.3.1. Proof of Theorem 3.5

Let $\{\infty,\kappa_d\}$ be as in the theorem statement. Proposition C.6 delivers that $\boldsymbol{\vartheta}\in\mathscr{C}(\{\infty,\kappa_d\})$ is equivalent to the condition that every subsequence d' of d has a subsequence d'' along which

$$\overline{\Phi}(\kappa_d-\|\boldsymbol{\theta}_d\|_\infty)\to 1 \quad \text{or} \quad \sum_{i=1}^{d}\overline{\Phi}\left(\kappa_d-|\theta_{i,d}|\right)\to\infty.$$

As a first step, we now show that this condition, and hence also $\boldsymbol{\vartheta}\in\mathscr{C}(\{\infty,\kappa_d\})$, is equivalent to the same condition but with $\mathfrak{c}_d=\sqrt{2\log(d)}-\frac{\log\log(d)}{2\sqrt{2\log(d)}}$ replacing κ_d.

Let d' be a subsequence of d. Lemma A.2 shows that $\kappa_d - \mathfrak{c}_d = u_d/\sqrt{2\log(d)}$ for a bounded sequence u_d. Therefore,

$$\limsup_{d'\to\infty} \overline{\Phi}(\kappa_{d'} - \|\boldsymbol{\theta}_{d'}\|_\infty) = 1 \quad \Leftrightarrow \quad \limsup_{d'\to\infty} \overline{\Phi}(\mathfrak{c}_{d'} - \|\boldsymbol{\theta}_{d'}\|_\infty) = 1, \tag{57}$$

and setting $\mathcal{I}_d := \{i \in \{1, \dots, d\} : \kappa_d - |\theta_{i,d}| \le 3\}$, it is obvious that

$$\limsup_{d'\to\infty} \sum_{i\in\mathcal{I}_{d'}} \overline{\Phi}\left(\kappa_{d'} - |\theta_{i,d'}|\right) = \infty \Leftrightarrow \limsup_{d'\to\infty} \sum_{i\in\mathcal{I}_{d'}} \overline{\Phi}\left(\mathfrak{c}_{d'} - |\theta_{i,d'}|\right) = \infty, \tag{58}$$

(as both divergences are equivalent to $|\mathcal{I}_d| \to \infty$ along a subsequence of d'). For $\mathcal{I}_d^c := \{1, \dots, d\} \setminus \mathcal{I}_d$ we now show that

$$\limsup_{d'\to\infty} \sum_{i\in\mathcal{I}_{d'}^c} \overline{\Phi}\left(\kappa_{d'} - |\theta_{i,d'}|\right) = \infty \Leftrightarrow \limsup_{d'\to\infty} \sum_{i\in\mathcal{I}_{d'}^c} \overline{\Phi}\left(\mathfrak{c}_{d'} - |\theta_{i,d'}|\right) = \infty. \tag{59}$$

To this end, let $d \ge 2$ be large enough so that

$$|u_d| \le \sqrt{2\log(d)}. \tag{60}$$

Furthermore, let x and y be in $(2, \infty)$ and such that $x - y = u_d/\sqrt{2\log(d)}$. We make the following two observations:

O1: If $\max(x, y) \ge \sqrt{2\log(d)}$, then, denoting $z := \sqrt{2\log(d)} - |u_d|/\sqrt{2\log(d)}$, we have

$$\overline{\Phi}(x) \vee \overline{\Phi}(y) \le \overline{\Phi}\left(z\right) \le e^{-z^2/2} \le d^{-1}e^{|u_d|},$$

where we used that $\overline{\Phi}(w) \le e^{-w^2/2}$ for $w \ge 0$.

O2: If $\max(x, y) \le \sqrt{2\log(d)})$, then $|x^2 - y^2| \le 2|u_d|$, and

$$\frac{x^{-1}}{y^{-1} - y^{-3}} = \left[(y + u_d/\sqrt{2\log(d)})(y^{-1} - y^{-3})\right]^{-1} \le \left[(1 - y^{-1})(1 - y^{-2})\right]^{-1} \le \frac{8}{3},$$

where we used (60) in the first (displayed) inequality. Now, (53) and boundedness of u_d shows that

$$0 < \frac{3}{8} e^{-\sup_{d\in\mathbb{N}} |u_d|} \le \frac{\overline{\Phi}(x)}{\overline{\Phi}(y)} \le \frac{8}{3} e^{\sup_{d\in\mathbb{N}} |u_d|} < \infty;$$

(by symmetry one only needs to check the upper bound).

Let $\mathcal{I}_{d,1}^c$ denote those indices i in $\mathcal{I}_d^c$ for which $\max(\kappa_d - |\theta_{i,d}|, \mathfrak{c}_d - |\theta_{i,d}|) \ge \sqrt{2\log(d)}$, and denote the set of remaining indices in $\mathcal{I}_d^c$ by $\mathcal{I}_{d,2}^c$. It follows from O1 together with $|\mathcal{I}_d^c| \le d$, that (to show the equivalence in (59)) it suffices to verify the equivalence with $\mathcal{I}_{d'}^c$ replaced by $\mathcal{I}_{d',2}^c$. But the latter equivalence follows from O2, which establishes (59). Combining (57), (58) and (59), we have shown that $\boldsymbol{\vartheta} \in \mathscr{C}(\{\infty, \kappa_d\})$ is equivalent to every subsequence d' of d having a subsequence d'' along which

$$\overline{\Phi}(\mathfrak{c}_d - \|\boldsymbol{\theta}_d\|_\infty) \to 1 \quad \text{or} \quad \sum_{i=1}^{d} \overline{\Phi}\left(\mathfrak{c}_d - |\theta_{i,d}|\right) \to \infty. \tag{61}$$

By Lemma A.2, the sequence of critical values $\mathfrak{c}_d$ satisfies $\{\infty, \mathfrak{c}_d\} \in \mathbb{T}_{\alpha^*}$ for some $\alpha^* \in (0,1)$. Hence, the just-derived equivalence (61) together with (the equivalence $(3 \Leftrightarrow 4)$) in Proposition C.6 applied to $\{\infty, \mathfrak{c}_d\}$ shows that

$$\boldsymbol{\vartheta} \in \mathscr{C}(\{\infty, \kappa_d\}) \Leftrightarrow \sum_{i=1}^{d} \frac{\overline{\Phi}\left(\mathfrak{c}_d - |\theta_{i,d}|\right)}{\Phi\left(\mathfrak{c}_d - |\theta_{i,d}|\right)} \to \infty. \tag{62}$$

Finally, note that by construction $\overline{\Phi}/g_\infty > 0$ is uniformly continuous on compact subsets of $\mathbb{R} \setminus \{1\}$ with positive left- and right-sided limits at 1, and $\overline{\Phi}(x)/g_\infty(x) \to 1/\sqrt{2\pi}$ as $x \to \infty$. Therefore, for every $z \in \mathbb{R}$ there exists a $C(z) \in (0,\infty)$ such that

$$\frac{\overline{\Phi}(x)}{g_\infty(x)} \in [C(z)^{-1}, C(z)] \qquad \text{for every } x \geq z. \tag{63}$$

Furthermore, a sequence x_d satisfies $g_\infty(x_d) \to \infty$ if and only if $x_d \to -\infty$. It follows that

$$\sum_{i=1}^{d} g_\infty\left(\mathfrak{c}_d - |\theta_{i,d}|\right) \to \infty$$

if and only if every subsequence d' of d permits a subsequence d'' along which $\mathfrak{c}_d - \|\boldsymbol{\theta}\|_\infty \to -\infty$ (equivalently $\overline{\Phi}(\mathfrak{c}_d - \|\boldsymbol{\theta}_d\|_\infty) \to 1$) or $\sum_{i=1}^d \overline{\Phi}(\mathfrak{c}_d - |\theta_{i,d}|) \to \infty$. In other words, we have shown that the condition in (61) is equivalent to $\sum_{i=1}^d g_\infty\left(\mathfrak{c}_d - |\theta_{i,d}|\right) \to \infty$, which concludes the proof.

A.3.2. *Proof of Theorem 3.7*

Proposition C.7 (applicable due to Remark C.3) and Remark C.4 deliver the result.

A.3.3. *Proof of Theorem 4.1*

It suffices to verify $\boldsymbol{\vartheta}^\dagger \in \mathscr{C}(\{p, \kappa_{d,p}\})$ and (27). Recall that $\boldsymbol{\theta}_d^\dagger := (\tau_d, \ldots, \tau_d, 0, \ldots, 0)$, where $\tau_d = \sqrt{2\log(d)/\log\log(d)}$, and that the number of non-zero entries of $\boldsymbol{\theta}_d$ is $\lceil \sqrt{d}/\log(d) \rceil$ (at least for d large enough).

1.: That $\boldsymbol{\vartheta}^\dagger \in \mathscr{C}(\{p, \kappa_{d,p}\})$ follows from Corollary 3.2, noting that eventually

$$\|\boldsymbol{\theta}_d\|_p^p/\sqrt{d} \geq 2^{p/2} \log^{p/2-1}(d)/(\log\log(d))^p \to \infty,$$

the divergence following from $p > 2$.

2.: That $\mathbb{P}(\|\boldsymbol{\theta}_d^\dagger + \boldsymbol{\varepsilon}_d\|_2 \geq \kappa_{d,2}) \to \alpha_2$ follows from Theorem C.4 (cf. also Remark C.1) and

$$\sum_{i=1}^{d} g_2(\theta_{i,d})/\sqrt{d} = \|\boldsymbol{\theta}_d^\dagger\|_2^2/\sqrt{d} \to 0.$$

3.: We now show that $\mathbb{P}(\|\boldsymbol{\theta}_d^\dagger + \boldsymbol{\varepsilon}_d\|_\infty \leq \kappa_{d,\infty}) \to 1 - \alpha_\infty$. Setting $k_d := \lceil \sqrt{d}/\log(d) \rceil$, write $\mathbb{P}(\|\boldsymbol{\theta}_d^\dagger + \boldsymbol{\varepsilon}_d\|_\infty \leq \kappa_{d,\infty})$ as

$$\mathbb{P}(|\varepsilon_1 + \tau_d| \leq \kappa_{d,\infty})^{k_d} \left[1 - 2\Phi(-\kappa_{d,\infty})\right]^{d-k_d}. \tag{64}$$

Concerning the second factor in (64), B_d, say, observe that

$$B_d = \left[1 - 2\Phi(-\kappa_{d,\infty})\right]^{d-k_d} = \frac{\left[1 - 2\Phi(-\kappa_{d,\infty})\right]^{d}}{\left[1 - 2\Phi(-\kappa_{d,\infty})\right]^{d\times\frac{k_d}{d}}} \to 1 - \alpha_\infty,$$

where we used that $\{\infty, \kappa_{d,\infty}\} \in \mathbb{T}_{\alpha_\infty}$ is equivalent to $[1-2\Phi(-\kappa_{d,\infty})]^d \to 1 - \alpha_\infty$, and that, as $k_d/d \to 0$, one also has $[1-2\Phi(-\kappa_{d,\infty})]^{d\times\frac{k_d}{d}} \to 1$. It remains to show that the first factor in (64), A_d, say, converges to 1. Note that $\mathbb{P}(|\varepsilon_1 + \tau_d| \leq \kappa_{d,\infty})$ can be written as

$$\Phi(\kappa_{d,\infty} - \tau_d) - \Phi(-\kappa_{d,\infty} - \tau_d) \geq 1 - 2\Phi(-\kappa_{d,\infty} + \tau_d) = 1 - 2\overline{\Phi}(\kappa_{d,\infty} - \tau_d). \tag{65}$$

By (56) (and since $\{\infty, \kappa_{d,\infty}\} \in \mathbb{T}_{\alpha_\infty}$ with $\alpha_\infty \in (0,1)$) we eventually have $\kappa_{d,\infty} - \tau_d \geq \sqrt{\log(d)}$. Thus, we can eventually apply Bernoulli's inequality (that is, $(1+x)^m \geq 1 + mx$ for $m \in \mathbb{N} \cup \{0\}$ and $x \geq -1$) and use (53) to conclude that

$$A_d \geq 1 - 2k_d\overline{\Phi}\left(\sqrt{\log(d)}\right) \geq 1 - 2k_d/\sqrt{d} \to 1.$$

A.3.4. *Proof of Corollary 4.2*

First consider the case where α_2 and α_∞ are both greater than 0. Then, we just note that (28) implies

$$\mathbb{E}\varphi_d(\boldsymbol{\theta}_d^\dagger + \boldsymbol{\varepsilon}_d) \leq \mathbb{P}(\|\boldsymbol{\theta}_d^\dagger + \boldsymbol{\varepsilon}_d\|_2 \geq \kappa_{d,2}) + \mathbb{P}(\|\boldsymbol{\theta}_d^\dagger + \boldsymbol{\varepsilon}_d\|_\infty \geq \kappa_{d,\infty}) \to \alpha_2 + \alpha_\infty < 1,$$

the convergence following from the last statement in Theorem 4.1. If any of the asymptotic sizes is 0, just pass to smaller critical values so that the asymptotic sizes are in $(0,1)$ and sum up to a number smaller than 1. Then, use the monotonicity of the rejection probabilities in the critical values and apply the already established statement.

A.3.5. *Proof of Theorem 5.1*

The theorem follows immediately from Theorem C.9 (cf. Remark C.1) together with Remarks C.3, C.4 and Theorem 3.5.

A.3.6. *Proof of Theorem 5.2*

This is an immediate consequence of Theorem C.10 (cf. Remark C.1), noting that by Anderson's theorem

$$\mathbb{E}(\psi_d(\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d)) - \mathbb{P}\left(\|\boldsymbol{\theta}_d + \boldsymbol{\varepsilon}_d\|_{p_j} \geq \kappa_d\right) \leq 1 - \mathbb{P}\left(\|\boldsymbol{\varepsilon}_d\|_{p_j} \geq \kappa_d\right) \leq 1 - \alpha + o(1),$$

which establishes (37) from Equation (C.54).

References

ANDERSON, T. W. (1955). The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. *Proc. Amer. Math. Soc.* **6** 170–176.

BIRNBAUM, A. (1955). Characterizations of complete classes of tests of some multiparametric hypotheses, with applications to likelihood ratio tests. *Ann. Math. Statist.* **26** 21–36.

BOGACHEV, L. (2006). Limit laws for norms of iid samples with Weibull tails. *J. Theoret. Probab.* **19** 849–873.

FAN, J., LIAO, Y. and YAO, J. (2015). Power enhancement in high-dimensional cross-sectional tests. *Econometrica* **83** 1497–1541.

FELLER, W. (1968). *An Introduction to Probability Theory and its Applications. Vol. I. Third edition*. New York: John Wiley & Sons.

FENG, L., JIANG, T., LIU, B. and XIONG, W. (2022). Max-sum tests for cross-sectional independence of high-dimensional panel data. *Ann. Statist.* **50** 1124 – 1143.

GIESSING, A. and FAN, J. (2020). Bootstrapping ℓ_p-statistics in high dimensions. *arXiv preprint arXiv:2006.13099*.

HE, Y., XU, G., WU, C. and PAN, W. (2021). Asymptotically independent U-statistics in high-dimensional testing. *Ann. Statist.* **49** 154–181.

INGSTER, Y. and SUSLINA, I. A. (2003). *Nonparametric Goodness-of-fit Testing under Gaussian Models*. New York: Springer.

JAMMALAMADAKA, S. R., MEINTANIS, S. and VERDEBOUT, T. (2020). On Sobolev tests of uniformity on the circle with an extension to the sphere. *Bernoulli* **26** 2226–2252.

JANSSEN, A. (2010). Limit laws for power sums and norms of i.i.d. samples. *Probab. Theory Related Fields* **146** 515–535.

KOCK, A. B. and PREINERSTORFER, D. (2019). Power in high-dimensional testing problems. *Econometrica* **87** 1055–1069.

KOCK, A. B. and PREINERSTORFER, D. (2021). Supplement to “Consistency of p-norm based tests in high dimensions: characterization, monotonicity, domination”.

LEADBETTER, M. R., LINDGREN, G. and ROOTZÉN, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer: New York.

PINELIS, I. (2010). Asymptotic efficiency of p-mean tests for means in high dimensions. *arXiv preprint arXiv:1006.0505*.

PINELIS, I. (2014). Schur 2-concavity properties of Gaussian measures, with applications to hypotheses testing. *J. Multivariate Anal.* **124** 384 - 397.

SCHLATHER, M. (2001). Limit distributions of norms of vectors of positive i.i.d. random variables. *Ann. Probab.* **29** 862–881.

STEIN, C. (1956). The admissibility of Hotelling’s T^2-test. *Ann. Math. Statist.* **27** 616–623.

WANG, H. J., MCKEAGUE, I. W. and QIAN, M. (2018). Testing for marginal linear effects in quantile regression. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **80** 433–452.

XU, G., LIN, L., WEI, P. and PAN, W. (2016). An adaptive two-sample test for high-dimensional means. *Biometrika* **103** 609–624.

YANG, Q. and PAN, G. (2017). Weighted statistic in detecting faint and sparse alternatives for high-dimensional covariance matrices. *J. Amer. Statist. Assoc* **112** 188–200.

YU, X., LI, D. and XUE, L. (2020). Fisher’s combined probability test for high-dimensional covariance matrices. *arXiv preprint arXiv:2006.00426*.

YU, X., LI, D., XUE, L. and LI, R. (2021). Power-enhanced simultaneous test of high-dimensional mean vectors and covariance matrices with application to gene-set testing. *arXiv preprint arXiv:2109.15287*.

ZHANG, Y., WANG, R. and SHAO, X. (2021). Adaptive inference for change points in high-dimensional data. *J. Amer. Statist. Assoc* forthcoming.