Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2018, Vol. 54, No. 2, 842–865
https://doi.org/10.1214/17-AIHP823

ANNALES DE L'INSTITUT HENRI POINCARÉ
PROBABILITÉS ET STATISTIQUES
www.imstat.org/aihp

Multidimensional two-component Gaussian mixtures detection

Béatrice Laurent[a], Clément Marteau[b] and Cathy Maugis-Rabusseau[a]

[a]*Institut de Mathématiques de Toulouse, INSA de Toulouse, Université de Toulouse, INSA de Toulouse, 135, avenue de Rangueil, 31077 Toulouse Cedex 4, France. E-mail: beatrice.laurent@insa-toulouse.fr; cathy.maugis@insa-toulouse.fr*
[b]*Univ Lyon, Université Claude Bernard Lyon 1, CNRS UMR 5208, Institut Camille Jordan, 43 blvd. du 11 novembre 1918, F-69622 Villeurbanne cedex, France. E-mail: marteau@math.univ-lyon1.fr*

Received 1 October 2015; revised 16 February 2017; accepted 22 February 2017

Abstract. Let $(X_1, \dots, X_n)$ be a d-dimensional i.i.d. sample from a distribution with density f. The problem of detection of a two-component mixture is considered. Our aim is to decide whether f is the density of a standard Gaussian random d-vector ($f = \phi_d$) against f is a two-component mixture: $f = (1-\varepsilon)\phi_d + \varepsilon\phi_d(\cdot - \mu)$ where (ε, μ) are unknown parameters. Optimal separation conditions on ε, μ, n and the dimension d are established, allowing to separate both hypotheses with prescribed errors. Several testing procedures are proposed and two alternative subsets are considered.

Résumé. Soit $(X_1, \dots, X_n)$ un n-échantillon d-dimensionnel dont la loi admet une densité f. Le problème de détection d'un mélange à deux composantes est étudié. Notre objectif est de déterminer si f est la densité de la loi gaussienne centrée réduite d-dimensionnelle ($f = \phi_d$) contre f est un mélange à deux composantes : $f = (1-\varepsilon)\phi_d + \varepsilon\phi_d(\cdot - \mu)$ où (ε, μ) sont des paramètres inconnus. Des conditions de séparation optimales sur ε, μ, n et la dimension d sont établies, permettant de séparer les deux hypothèses à erreurs fixées. Plusieurs procédures de test sont proposées et deux sous-ensembles d'alternatives sont considérés.

MSC: Primary 62H15; secondary 62G30

Keywords: Gaussian mixtures; Non-asymptotic testing procedure; Separation rates

1. Introduction

Let $\underline{X} = (X_1, \dots, X_n)$ be an i.i.d. n-sample, where for all $i \in \{1, \dots, n\}$, X_i corresponds to a d-dimensional random vector, whose distribution admits a density f w.r.t. the Lebesgue measure on $\mathbb{R}^d$. In the following, we denote by $\phi_d(\cdot)$ the density function of the standard Gaussian distribution $\mathcal{N}_d(0_d, I_d)$ on $\mathbb{R}^d$. Our aim is to test

$$H_0 : f = \phi_d \quad \text{against } H_1 : f \in \mathcal{F}, \tag{1}$$

where

$$\mathcal{F} = \left\{ f_{(\varepsilon,\mu)} : x \in \mathbb{R}^d \mapsto (1-\varepsilon)\phi_d(x) + \varepsilon\phi_d(x-\mu); \varepsilon \in]0, 1[, \mu \in \mathbb{R}^d \right\}$$

is the set of two-component Gaussian mixtures on $\mathbb{R}^d$. Mixture models are at the core of several studies and provide a powerful paradigm that allows to model several practical phenomena. We refer to [20] for an extended introduction to this topic.

The particular case of a two-component mixture is sometimes referred as a contamination model. In some sense, a proportion ε of the sample is driven from a (Gaussian) distribution centered in μ while the remaining part of the data is centered. In this context, the testing problem (1) amounts to the detection of a plausible contamination inside the

data at hand w.r.t. the null distribution. We refer for instance to [12] for practical motivations regarding this problem. We stress that Gaussian mixture is at the core of our contribution since it provides a benchmark model for several practical applications. However, the results proposed in this paper could be certainly extended to a wide range of alternative distributions.

In a unidimensional setting ($d = 1$), the testing problem (1) has been widely considered in the literature in the last two decades. A large attention has been payed to methods based on the likelihood ratio, see e.g. [2,11] or [14]. Concerning the construction of optimal separation conditions on the parameters (ε, μ), we can mention the seminal contribution of Ingster [16]. These conditions have been reached by the higher-criticism procedure proposed by Donoho and Jin [12] in a specific *sparse* context, i.e. when $\varepsilon \ll 1/\sqrt{n}$ as $n \to +\infty$. Then, several extensions of this contribution have been proposed in an extended context: we mention for instance [9] for a study including confidence sets and the *dense* setting ($\varepsilon \gg 1/\sqrt{n}$ as $n \to +\infty$), [8] for heterogeneous and heteroscedastic mixtures, or [10] where general distributions and separation conditions have been investigated. In a slightly different spirit, a procedure based on the order statistics and non-asymptotic investigations on the testing problem (1) have been proposed in [18].

In the contributions mentioned above, only unidimensional distributions are considered. In a different setting (signal detection), multidimensional problems have been at the core of recent investigations. We mention e.g. [1] or [7] among others. In a recent paper, Verzelen and Arias-Castro [21] address the problem of testing normality in a multidimensional framework. They consider two-component Gaussian mixture alternatives where the proportions are fixed and the difference in means are sparse. However, up to our knowledge, the multidimensional testing problem as displayed in (1) has never been studied so far. We stress that in our setting, the proportion ε is allowed to depend on the number of observations n. The present paper proposes a first attempt in this context.

The testing problem, as formalized in (1) does not allow to guarantee a possible separation between H_0 and H_1 with prescribed Type I and Type II errors. Indeed, we can construct mixture distributions arbitrarily close (in a sense that should be made precise) to the Gaussian law. To this end, we will restrict our analysis to the mixtures $f_{(\varepsilon,\mu)} \in \mathcal{F}$ satisfying $\varepsilon\|\mu\|_\square \geq \rho$, where $\|\cdot\|_\square$ will alternatively denote the l_2 and l_∞ norms, and $\rho > 0$ a given radius. In each case, our aim is to investigate the slowest possible value of the radius ρ for which both hypotheses can be separated. In this multidimensional setting, the definition of *dense* and *sparse* regime is more involved since the dimension d of the problem has a real influence on the detection problem. In this context, we will provide a sharp description of the optimal separation radius ρ in a case which could be considered as *dense* for both norms. On the other hand, we will provide some attempts in the *sparse* regime.

The paper is organized as follows. Section 2 is devoted to the l_2-norm. A lower bound is proposed in Section 2.1. Two computationally tractable testing procedures ($\Psi_{1,\alpha}$ and $\Psi_{2,\alpha}$) are studied in Section 2.2 and an upper bound for the aggregation of these both tests is established. One computationally intractable test procedure ($\Psi_{3,\alpha}$) is described in Section 2.3 and an upper bound (improving the previous result) is obtained for the aggregation of the tests $\Psi_{1,\alpha}$ and $\Psi_{3,\alpha}$. Section 3 is devoted to the l_∞-norm. A lower bound is proposed in Section 3.1. The performances of two different testing procedures ($\Psi_{4,\alpha}$ and $\Psi_{5,\alpha}$) are investigated and an upper bound for the aggregation of these both tests is established in Section 3.2. A short discussion gathering remaining open problems and possible outcomes is presented in Section 4. The proofs of lower and upper bounds are presented in Sections 5 and 6 respectively. Some useful lemmas are gathered in Appendix A, while Appendix B contains some technical results for unidimensional two-component Gaussian mixtures detection.

All along the paper, we use the following notations. For any density g on $\mathbb{R}^d$, we denote respectively by $\mathbb{P}_g$ and $\mathbb{E}_g$ the probability and expectation under the assumption that the common density of each X_i in the i.i.d. sample $\underline{X} = (X_1, \ldots, X_n)$ is g. In the particular case where the $X_1, \ldots, X_n$ are i.i.d. with common density ϕ_d, which is associated to the null hypothesis H_0, we write $\mathbb{P}_0 := \mathbb{P}_{\phi_d}$ and $\mathbb{E}_0 := \mathbb{E}_{\phi_d}$. A testing procedure Ψ denotes a measurable function of the sample $\underline{X}$, having values in $\{0, 1\}$. By convention, we reject (resp. do not reject) H_0 if $\Psi = 1$ (resp. $\Psi = 0$). Given $\alpha \in]0, 1[$, the test Ψ is said to be of level α if $\mathbb{P}_0(\Psi = 1) \leq \alpha$. In such a case, we write $\Psi = \Psi_\alpha$. For any vector $\mu \in \mathbb{R}^d$, we set $\|\mu\| = (\sum_{j=1}^d \mu_j^2)^{1/2}$ and $\|\mu\|_\infty = \max_{1\leq j\leq d} |\mu_j|$.

2. Detection boundary for the l_2-norm

2.1. *Lower bound for the l_2-norm in the dense regime*

The non-asymptotic minimax separation rates have been introduced in [3]. Let us recall the main definitions. Given $\beta \in]0, 1[$, the class of alternatives $\mathcal{F}$ and a level-α test Ψ_α, we define the uniform separation $\rho(\Psi_\alpha, \mathcal{F}, \beta)$ of Ψ_α with

respect to the l_2-norm over the class $\mathcal{F}$ as the smallest positive number ρ such that the test has a second kind error at most equal to β for all alternatives $f_{(\varepsilon,\mu)}$ in $\mathcal{F}$ such that $\varepsilon\|\mu\| \geq \rho$. More precisely,

$$\rho(\Psi_\alpha, \mathcal{F}, \beta) = \inf\Big\{\rho > 0;\ \sup_{\substack{f\in\mathcal{F}\\ \varepsilon\|\mu\|\geq\rho}} \mathbb{P}_f(\Psi_\alpha = 0) \leq \beta\Big\}.$$

Then, the (α, β)-minimax separation rate over $\mathcal{F}$ is defined as

$$\underline{\rho}(\mathcal{F}, \alpha, \beta) = \inf_{\Psi_\alpha} \rho(\Psi_\alpha, \mathcal{F}, \beta),$$

where the infimum is taken over all level-α tests Ψ_α.

Theorem 1 proposes a lower bound for the minimax separation rate $\underline{\rho}(\mathcal{F}, \alpha, \beta)$ under the assumption that $\varepsilon \geq d^{1/4}/\sqrt{n}$, which is the case that we call the *dense regime*.

Theorem 1. *Assume that* $\varepsilon \geq d^{1/4}/\sqrt{n}$. *Let* $\alpha, \beta \in]0, 1[$ *such that* $\alpha + \beta < 0.29$. *Define*

$$\rho^{\#} = 0.4\frac{d^{1/4}}{\sqrt{n}}.$$

Then, if $\rho < \rho^{\#}$,

$$\inf_{\Psi_\alpha} \sup_{\substack{f\in\mathcal{F}\\ \varepsilon\|\mu\|\geq\rho}} \mathbb{P}_f(\Psi_\alpha = 0) > \beta, \tag{2}$$

where the infimum is taken over all level-α *tests. In particular, this implies that*

$$\underline{\rho}(\mathcal{F}, \alpha, \beta) \geq \rho^{\#}.$$

Equation (2) indicates that the hypotheses H_0 and H_1 cannot be separated with prescribed first and second kind errors α and β following the value of the terms ε, $\|\mu\|$, d and n. In particular, for any level-α testing procedure, one can find a distribution $f \in \mathcal{F}$ such that $\varepsilon\|\mu\| \geq \rho^{\#}$ and $\mathbb{P}_f(\Psi_\alpha = 0) > \beta$. This result is obtained thanks to the assumption $\alpha + \beta < 0.29$. This assumption is essentially technical and could be removed with additional technical algebra.

The condition $\varepsilon\|\mu\| \geq Cd^{1/4}/\sqrt{n}$ for some constant $C > 0$ is quite informative. For a given $\varepsilon \geq d^{1/4}/\sqrt{n}$, we can specify how the 'energy' $\|\mu\|$ should be large if one expects to detect a potential contamination in the sample. It is worth pointing out that Theorem 1 precisely quantifies the role played by the dimension d of the problem at hand. We will see in Section 2.2 that this lower bound is optimal, up to some constant.

The main ingredient for the proof (displayed in Section 5) is the construction of particular distributions for which the separation of both hypotheses H_0 and H_1 will be impossible with a prescribed level β.

2.2. *Computationally tractable upper bounds for the l_2-norm*

In Section 2.1, we have proposed lower bounds on the separation rates for the testing problem (1) in the *dense regime*. In particular, we have proved that in some specific cases, related to the value of the parameters (ε, μ), testing is impossible, i.e. every level-α test will be associated to a second kind error greater than a prescribed level β.

The aim of this section is to complete this discussion with upper bounds on the separation rates. We propose two different testing procedures and investigate their related performances. In particular, we prove that these procedures reach the lower bounds presented in Theorem 1.

The first procedure is a very simple test based on the fluctuations of the empirical mean of the data. Intuitively, $\mathbb{E}_f[X] = \varepsilon\mu$ for all random vectors having density $f \in \mathcal{F}$, while $\mathbb{E}_0[X] = 0$ under H_0. In particular, if the empirical mean of the sample has a large norm, there is a chance that the data have been driven w.r.t. a density f that belongs to $\mathcal{F}$.

More precisely, given $\underline{X} = (X_1, \dots, X_n)$, set $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Let $t_{d,\alpha}$ denote the $(1-\alpha)$ quantile of a chi-square distribution with d degrees of freedom and define the test $\Psi_{1,\alpha}$ as

$$\Psi_{1,\alpha} = \mathbb{1}_{\{\|\sqrt{n}\bar{X}_n\|^2 > t_{d,\alpha}\}}. \tag{3}$$

The following theorem investigates the performances of this test.

Theorem 2. *Let $\alpha, \beta \in]0,1[$ be fixed. We assume that $n\varepsilon \geq \frac{8}{\beta}$. Then, the testing procedure $\Psi_{1,\alpha}$ introduced in (3) is of level α. Moreover, there exists a positive constant $C(\alpha,\beta)$, only depending on α and β, such that, for all $f = f_{(\varepsilon,\mu)} \in \mathcal{F}$ for which*

$$\varepsilon^2\|\mu\|^2 \geq C(\alpha,\beta)\frac{\sqrt{d}}{n},$$
$$\mathbb{P}_f(\Psi_{1,\alpha} = 0) \leq \beta.$$

The above result indicates that the test $\Psi_{1,\alpha}$ is powerful as soon as $f \in \mathcal{F}$ with $\varepsilon\|\mu\| \geq C(\alpha,\beta)d^{1/4}/\sqrt{n}$. According to the lower bound displayed in Theorem 1, it appears that in the so-called *dense regime*, i.e. when $\varepsilon \geq d^{1/4}/\sqrt{n}$, the minimax detection frontier is of order $d^{1/4}/\sqrt{n}$ up to a constant, i.e. there exist $\mathcal{C}_-$ and $\mathcal{C}_+$ such that

- the hypotheses H_0 and H_1 cannot be separated if $\varepsilon\|\mu\| \leq \mathcal{C}_- d^{1/4}/\sqrt{n}$,
- there exists a level-α powerful test as soon as $\varepsilon\|\mu\| \geq \mathcal{C}_+ d^{1/4}/\sqrt{n}$.

We stress that we do not investigate the value of the optimal constant associated to this separation problem ($\mathcal{C}_-$ and $\mathcal{C}_+$ do not match). Such a study indeed requires advanced asymptotic tools [see e.g. [15]] and is outside the scope of the paper.

The procedure proposed in (3) is optimal in the *dense regime*. We will now introduce another computationally tractable procedure that improves the performances of the previous test in the *sparse regime*, namely when $\varepsilon < d^{1/4}/\sqrt{n}$. The test statistics is defined as

$$\Psi_{2,\alpha} = \max_{1\leq i\leq n} \mathbb{1}_{\{\|X_i\|^2 > t_{d,\alpha/n}\}}, \tag{4}$$

where $t_{d,\alpha/n}$ denotes the $(1-\alpha/n)$ quantile of the chi-square distribution with d degrees of freedom. The performances of this procedure are given in Theorem 3.

Theorem 3. *Let $\alpha, \beta \in]0,1[$ be fixed. We assume that $n\varepsilon \geq \frac{8}{\beta}$. Then, the testing procedure $\Psi_{2,\alpha}$ introduced in (4) is of level α. Moreover, there exists a positive constant $C(\alpha,\beta)$ only depending on α and β such that for all $f = f_{(\varepsilon,\mu)} \in \mathcal{F}$ for which*

$$\varepsilon^2\|\mu\|^2 \geq C(\alpha,\beta)\varepsilon^2\big[\sqrt{d\ln(n)} + \ln(n)\big],$$
$$\mathbb{P}_f(\Psi_{2,\alpha} = 0) \leq \beta.$$

We can easily aggregate the tests $\Psi_{1,\alpha}$ and $\Psi_{2,\alpha}$ by considering the test function $\Psi_{1,\alpha/2} \vee \Psi_{2,\alpha/2}$. Noticing that

$$\mathbb{P}_0(\Psi_{1,\alpha/2} \vee \Psi_{2,\alpha/2} = 1) \leq \mathbb{P}_0(\Psi_{1,\alpha/2} = 1) + \mathbb{P}_0(\Psi_{2,\alpha/2} = 1) \leq \alpha,$$

this leads to a level-α test. Moreover,

$$\mathbb{P}_f(\Psi_{1,\alpha/2} \vee \Psi_{2,\alpha/2} = 0) \leq \inf\big\{\mathbb{P}_f(\Psi_{1,\alpha/2} = 0), \mathbb{P}_f(\Psi_{2,\alpha/2} = 0)\big\},$$

hence the second kind error of the aggregated test is controlled (up to constants since α has been replaced by $\alpha/2$) by the smallest second kind error of the two tests. This leads to the following result.

Theorem 4. *Let $\alpha, \beta \in]0,1[$ be fixed. Let $\Psi_{1,\alpha}$ and $\Psi_{2,\alpha}$ be the both tests defined in Equations (3) and (4) respectively. There exists a positive constant $C(\alpha,\beta)$ only depending on α, β such that, if $n\varepsilon \geq \frac{8}{\beta}$, for all $f = f_{(\varepsilon,\mu)} \in \mathcal{F}$ which fulfills*

$$\varepsilon^2 \|\mu\|^2 \geq C(\alpha,\beta)\left[\left(\frac{\sqrt{d}}{n}\right) \wedge \left\{\varepsilon^2\left(\sqrt{d\ln(n)} + \ln(n)\right)\right\}\right], \tag{5}$$

we have

$$\mathbb{P}_f(\Psi_{1,\frac{\alpha}{2}} \vee \Psi_{2,\frac{\alpha}{2}} = 0) \leq \beta.$$

It is important to compute the right hand term of Inequality (5) to see how this result improves the separation condition established in Theorem 2. We define

$$\rho^2_{n,d,\varepsilon} = \left(\frac{\sqrt{d}}{n}\right) \wedge \left\{\varepsilon^2\left(\sqrt{d\ln(n)} + \ln(n)\right)\right\}.$$

The separation conditions are summarized as follows:

- If $d \leq \ln(n)$, then
  - If $\varepsilon \leq d^{1/4}/\sqrt{n\ln(n)}$, then $\rho^2_{n,d,\varepsilon} \leq 2\varepsilon^2\ln(n)$.
  - If $\varepsilon \geq d^{1/4}/\sqrt{n\ln(n)}$, then $\rho^2_{n,d,\varepsilon} \leq \sqrt{d}/n$.
- If $d \geq \ln(n)$, then
  - If $\varepsilon \leq 1/(\sqrt{n}(\ln(n))^{1/4})$, then $\rho^2_{n,d,\varepsilon} \leq 2\varepsilon^2\sqrt{d\ln(n)}$.
  - If $\varepsilon \geq 1/(\sqrt{n}(\ln(n))^{1/4})$, then $\rho^2_{n,d,\varepsilon} \leq \sqrt{d}/n$.

We therefore see that the separation rate of the aggregated test is smaller than the one of the test $\Psi_{1,\alpha}$ in the case where $d \leq \ln(n)$ and $\varepsilon \leq d^{1/4}/\sqrt{n\ln(n)}$, and in the case where $d \geq \ln(n)$ and $\varepsilon \leq 1/(\sqrt{n}(\ln(n))^{1/4})$. Unfortunately we do not know if these results are optimal since we did not manage to get lower bounds for the l_2-norm in the *sparse regime*, namely when $\varepsilon < d^{1/4}/\sqrt{n}$.

2.3. *Computationally intractable upper bounds for the l_2-norm*

During the revision process, a referee suggested to consider an alternative testing procedure $\Psi_{3,\alpha}$ defined as follows

$$\Psi_{3,\alpha} = \sup_{U\in\mathcal{U}} \mathbb{1}_{\{T_U > t_{n,d,|U|,\alpha}\}}, \tag{6}$$

where $\mathcal{U}$ denotes the set of the nonempty subsets of $\{1,\ldots,n\}$, $|U|$ denotes the cardinality of U,

$$T_U = \frac{1}{|U|}\left\|\sum_{i\in U} X_i\right\|^2,$$

$t_{n,d,k,\alpha} = d + 2\sqrt{dx_{n,k,\alpha}} + 2x_{n,k,\alpha}$ and $x_{n,k,\alpha} = k\ln(en/k) + \ln(n/\alpha)$. The statistical performances are presented in the following theorem, whose proof is postponed to Section 6.3.

Theorem 5. *Let $\alpha, \beta \in]0,1[$ be fixed. Let $\Psi_{1,\alpha}$ and $\Psi_{3,\alpha}$ be the both tests defined in Equations (3) and (6) respectively. There exists a positive constant $C(\alpha,\beta)$ only depending on α, β such that, for all $f = f_{(\varepsilon,\mu)} \in \mathcal{F}$ which fulfills $n\varepsilon \geq \frac{8}{\beta}$ and*

$$\varepsilon^2\|\mu\|^2 \geq C(\alpha,\beta)\left[\left(\frac{\sqrt{d}}{n}\right) \wedge \left\{\varepsilon^2\ln\left(\frac{1}{\varepsilon}\right) + \varepsilon^{3/2}\sqrt{\frac{d}{n}\ln\left(\frac{1}{\varepsilon}\right)}\right\}\right], \tag{7}$$

we have

$$\mathbb{P}_f(\Psi_{1,\frac{\alpha}{2}} \vee \Psi_{3,\frac{\alpha}{2}} = 0) \leq \beta.$$

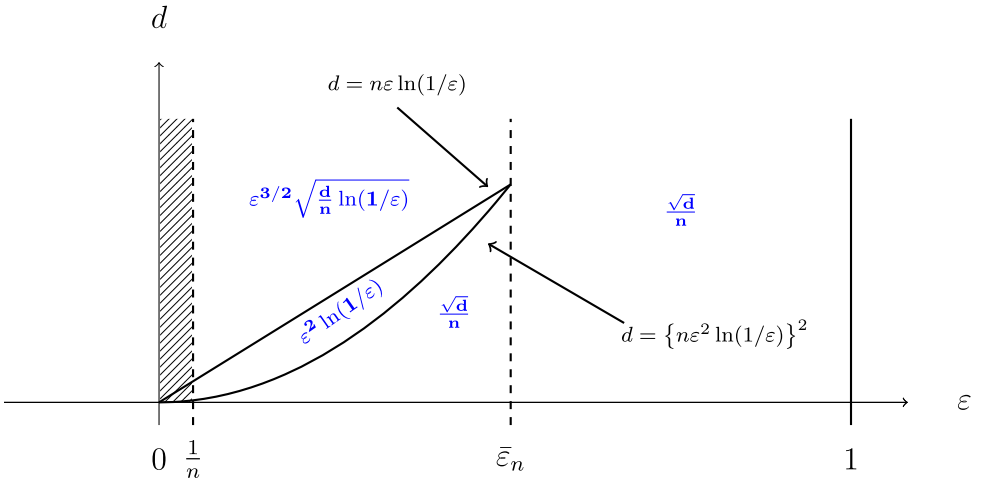

Fig. 1. Summary of the separation condition on $\varepsilon^2\|\mu\|^2$ for the test $\Psi_{1,\alpha/2} \vee \Psi_{3,\alpha/2}$, where $\bar{\varepsilon}_n = \inf\{\varepsilon \in]0,1[; n\varepsilon^3 \ln(1/\varepsilon) \geq 1\}$.

The optimal set $U \in \mathcal{U}$ that allows to derive the separation condition has a cardinal of order εn. A first remark concerning this result is that it improves the separation condition established in Theorem 4 since $\varepsilon \geq 1/n$. The separation conditions are summarized in Figure 1, the computations are detailed at the end of the proof of Theorem 5. However, it can be noticed that computing this testing procedure is probably NP-hard, contrary to the test $\Psi_{1,\alpha/2} \vee \Psi_{2,\alpha/2}$. This may explain the difference of performances between both approaches [we refer to [4,5] for an extended discussion in a different setting]. Unfortunately, we did not succeed in the construction of appropriate lower bounds. Hence, determining the minimax separation condition when $\varepsilon < d^{1/4}/\sqrt{n}$ with computationally tractable or not testing procedures remains an open problem.

3. Detection boundary for the l_∞-norm

3.1. *Lower bound for the l_∞-norm*

As in Section 2.1, we consider the (α, β)-minimax separation rate over $\mathcal{F}$ with respect to the l_∞-norm defined as

$$\underline{\rho}_\infty(\mathcal{F}, \alpha, \beta) = \inf_{\Psi_\alpha} \rho_\infty(\Psi_\alpha, \mathcal{F}, \beta),$$

where the infimum is taken over all level-α tests Ψ_α and

$$\rho_\infty(\Psi_\alpha, \mathcal{F}, \beta) = \inf\Big\{\rho > 0; \sup_{\substack{f \in \mathcal{F} \\ \varepsilon\|\mu\|_\infty \geq \rho}} \mathbb{P}_f(\Psi_\alpha = 0) \leq \beta\Big\}.$$

Theorem 6 provides a lower bound for the minimax separation rate in this context, the proof is postponed to Section 5.

Theorem 6. *Let $\alpha, \beta \in]0,1[$ such that $\alpha + \beta < 1$. Let $\eta(\alpha, \beta) = 2(1 - \alpha - \beta)$. Define*

$$\rho^\star = \varepsilon\sqrt{\ln\Big[1 + \frac{1}{n\varepsilon^2}\ln\big(1 + d\eta(\alpha,\beta)^2\big)\Big]}.$$

Then, if $\rho \le \rho^\star$,

$$\inf_{\Psi_\alpha} \sup_{\substack{f \in \mathcal{F} \\ \varepsilon \|\mu\|_\infty \ge \rho}} \mathbb{P}_f(\Psi_\alpha = 0) > \beta, \tag{8}$$

where the infimum is taken over all level-α *tests. In particular, this implies that*

$$\underline{\rho}_\infty(\mathcal{F}, \alpha, \beta) \ge \rho^\star.$$

As for the l_2-norm, two main different regimes can be distinguished from this lower bound:

- Case 1: If $n\varepsilon^2 \gg \ln(d)$, or if $c_1 \ln(d) \le n\varepsilon^2 \le c_2 \ln(d)$ for $0 < c_1 < c_2$, then $\rho^\star \sim \sqrt{\ln(d)/n}$.
- Case 2: If $n\varepsilon^2 \ll \ln(d)$, then $\rho^\star \sim \varepsilon\sqrt{\ln(\ln(d)/n\varepsilon^2)}$.

By analogy with the discussion conducted in the previous section, the first case can be considered as the *dense* regime; this case allows in particular to detect bounded contamination, i.e. contamination for which $\|\mu\|_\infty < \infty$. The control in Case 1 provides a similar separation condition compared to the l_2 case, except that the dependency with respect to d: the quantity $d^{1/4}$ is replaced here by $\sqrt{\ln(d)}$.

On the other hand, in the case where $n\varepsilon^2 \ll \ln(d)$, which can be considered as *sparse* regime in this l_∞ analysis, the separation condition is of the form $\varepsilon\sqrt{\ln(\ln(d)/n\varepsilon^2)}$. The optimality of these bounds are discussed in the next section.

3.2. *Testing procedures for the* l_∞*-norm*

In this section, we consider two testing procedures for the l_∞-norm. Both consist of applying a testing procedure on each canonical direction in order to detect a possible contamination.

For each $i \in \{1, \dots, n\}$, we denote $X_i = (X_{ij})_{1 \le j \le d}$. For a given $j \in \{1, \dots, d\}$, we can remark that $(X_{ij})_{i=1,\dots,n}$ is a unidimensional sample, distributed from

- the standard Gaussian distribution $\mathcal{N}(0, 1)$ under H_0,
- the unidimensional mixture $(1 - \varepsilon)\phi_1(\cdot) + \varepsilon\phi_1(\cdot - \mu_j)$ under H_1.

First, we consider the following testing procedure:

$$\Psi_{4,\alpha} = \max_{j=1,\dots,d} \mathbb{1}_{\{n\bar{X}_j^2 > t_{1,\frac{\alpha}{d}}\}}, \tag{9}$$

where $\bar{X}_j = \frac{1}{n}\sum_{i=1}^n X_{ij}$ and $t_{1,\frac{\alpha}{d}}$ is the $(1 - \frac{\alpha}{d})$-quantile of a chi-square distribution with one degree of freedom.

Theorem 7. *Let* $\alpha, \beta \in]0, 1[$ *be fixed. We assume that* $n\varepsilon \ge \frac{8}{\beta}$. *Then, the testing procedure* $\Psi_{4,\alpha}$ *introduced in* (9) *is of level* α. *Moreover, there exists a positive constant* $C(\alpha, \beta)$ *only depending on* α *and* β *such that for all* $f = f_{(\varepsilon,\mu)} \in \mathcal{F}$ *for which*

$$\varepsilon\|\mu\|_\infty \ge C(\alpha, \beta)\sqrt{\frac{\ln(d)}{n}},$$
$$\mathbb{P}_f(\Psi_{4,\alpha} = 0) \le \beta.$$

The second testing procedure is defined by

$$\Psi_{5,\alpha} = \max_{j=1,\dots,d} \mathbb{1}_{\{\max_{i=1,\dots,n} X_{ij}^2 > t_{1,\frac{\alpha}{dn}}\}}, \tag{10}$$

where $t_{1,\frac{\alpha}{dn}}$ is the $(1 - \frac{\alpha}{dn})$-quantile of a chi-square distribution with one degree of freedom.

Theorem 8. *Let $\alpha, \beta \in]0,1[$ be fixed. We assume that $n\varepsilon \geq \frac{8}{\beta}$. Then, the testing procedure $\Psi_{5,\alpha}$ introduced in (10) is of level α. Moreover, there exists a positive constant $C(\alpha,\beta)$ only depending on α and β such that for all $f = f_{(\varepsilon,\mu)} \in \mathcal{F}$ for which*

$$\|\mu\|_\infty \geq C(\alpha,\beta)\sqrt{\ln(dn)},$$
$$\mathbb{P}_f(\Psi_{5,\alpha} = 0) \leq \beta.$$

As explained in Section 2.2, we can easily aggregate the tests $\Psi_{4,\alpha}$ and $\Psi_{5,\alpha}$ by considering the test function $\Psi_{4,\alpha/2} \vee \Psi_{5,\alpha/2}$. This leads to the following result.

Theorem 9. *Let $\alpha, \beta \in]0,1[$ be fixed. Let $\Psi_{4,\alpha}$ and $\Psi_{5,\alpha}$ be the tests defined in Equations (9) and (10) respectively. There exists a positive constant $C(\alpha,\beta)$ only depending on α, β such that, if $n\varepsilon \geq \frac{8}{\beta}$, for all $f = f_{(\varepsilon,\mu)} \in \mathcal{F}$ which fulfills*

$$\varepsilon\|\mu\|_\infty \geq C(\alpha,\beta)\left[\sqrt{\frac{\ln(d)}{n}} \wedge \varepsilon\sqrt{\ln(dn)}\right], \tag{11}$$

we have

$$\mathbb{P}_f(\Psi_{4,\frac{\alpha}{2}} \vee \Psi_{5,\frac{\alpha}{2}} = 0) \leq \beta.$$

Concerning the upper bound for the separation rate established in Theorem 9, we obtain the following results:

- If $n\varepsilon^2 \geq \ln(d)/\ln(dn)$, the right hand term in Inequality (11) is of order $\sqrt{\ln(d)/n}$.
- If $n\varepsilon^2 \leq \ln(d)/\ln(dn)$, the right hand term in Inequality (11) is of order $\varepsilon\sqrt{\ln(dn)}$.

Let us now compare these upper bounds with the lower bounds obtained in Theorem 6. We see that in the case that we have denoted Case 1, or *dense case* where either $n\varepsilon^2 \gg \ln(d)$, or $c_1\ln(d) \leq n\varepsilon^2 \leq c_2\ln(d)$, the upper and lower bounds coincide and our results are optimal. In the other case, a gap remains between lower and upper bounds, which provides an open problem for this testing problem.

4. Discussions

In this paper, we have addressed the detection problem of a mixture distribution as formalized in (1). The alternative involved an energy condition through the expression $\varepsilon\|\mu\|_\square \geq \rho$ where $\|\cdot\|_\square$ has alternatively denoted the l_2 and l_∞ norms, and ρ a minimal value for which the hypotheses H_0 and H_1 can be separated with prescribed levels.

The results presented in Sections 2 and 3 provide a first attempt toward the description of optimal values for the quantity ρ. In the so-called *dense* regimes, $\varepsilon \gtrsim d^{1/4}/\sqrt{n}$ and $\varepsilon \gtrsim \sqrt{\ln(d)/n}$ for the l_2 and l_∞ norms respectively, Theorems 1, 2, 6 and 9 provide a precise characterization of the separation radius ρ. On the other hand, the sparse regimes seem to be more involved. In particular, we did not provide a sharp characterization of the dependency with respect to the dimension d for this detection problem. By the way, a deeper analysis will require discussions on computability conditions that are outside the initial purpose of this paper.

Several additional investigations could be driven in this setting, among them: considering more general benchmark distributions (i.e. different from the standard Gaussian distribution), heteroscedastic mixtures or taking into account some uncertainty on the reference distribution. All these questions are outside the scope of the paper but could be at the core of future contributions.

5. Proof of Theorems 1 and 6

For the sake of convenience, we introduce the subset $\mathcal{F}[\rho]$ which corresponds to

$$\mathcal{F}_2[\rho] = \left\{f \in \mathcal{F};\, \varepsilon\|\mu\| \geq \rho\right\}$$

in the first proof, and

$$\mathcal{F}_\infty[\rho] = \big\{ f \in \mathcal{F}; \varepsilon \|\mu\|_\infty \geq \rho \big\}$$

in the second proof, for any given radius $\rho > 0$. Following [3] or [15], we will use a Bayesian argument in order to bound the minimax separation radius in the two contexts. Thus, we consider a subset $\{g_\omega; \omega \in \Omega\}$ of $\mathcal{F}[\rho]$ which will be specified for each proof later. Then,

$$\sup_{f \in \mathcal{F}[\rho]} \mathbb{P}_f(\Psi_\alpha = 0) \geq \mathbb{P}_{g_\omega}(\Psi_\alpha = 0), \quad \forall \omega \in \Omega.$$

Denoting the uniform probability measure π on the finite set Ω, we have

$$\sup_{f \in \mathcal{F}[\rho]} \mathbb{P}_f(\Psi_\alpha = 0) \geq \int_\Omega \mathbb{P}_{g_\omega}(\Psi_\alpha = 0)\, d\pi(\omega) := \mathbb{P}_\pi(\Psi_\alpha = 0). \tag{12}$$

Using (12) and similar computations as in [3] or [15], we obtain

$$\begin{aligned} \inf_{\Psi_\alpha} \sup_{f \in \mathcal{F}[\rho]} \mathbb{P}_f(\Psi_\alpha = 0) &\geq \inf_{\Psi_\alpha} \mathbb{P}_\pi(\Psi_\alpha = 0) \\ &\geq 1 - \alpha - \frac{1}{2}\sqrt{\mathbb{E}_0\big[L_\pi^2(\underline{X})\big] - 1}, \end{aligned}$$

where $L_\pi(\underline{X}) = \frac{d\mathbb{P}_\pi}{d\mathbb{P}_0}(\underline{X})$ is the likelihood ratio. In particular, if we can ensure that

$$\mathbb{E}_0\big[L_\pi^2(\underline{X})\big] < 1 + \eta(\alpha, \beta)^2,$$

where $\eta(\alpha, \beta) = 2(1 - \alpha - \beta)$ for all $\alpha, \beta \in]0, 1[$, then

$$\inf_{\Psi_\alpha} \sup_{f \in \mathcal{F}[\rho]} \mathbb{P}_f(\Psi_\alpha = 0) > 1 - \alpha - \frac{1}{2}\eta(\alpha, \beta) = \beta.$$

In the two following proofs displayed below, we will specify the subset $\{g_\omega; \omega \in \Omega\}$ and propose an upper bound for the term $\mathbb{E}_0[L_\pi^2(\underline{X})]$.

5.1. *Proof of Theorem 1*

In this proof, recall that

$$\mathcal{F}[\rho] = \big\{ f \in \mathcal{F}; \varepsilon \|\mu\| \geq \rho \big\},$$

for any given radius $\rho > 0$.

We now consider $r > 0$ and $\varepsilon \in]0, 1[$ such that $\varepsilon r = \rho$. In this context, we choose $\Omega = \{-1, 1\}^d$ and

$$\forall \omega \in \Omega, \quad g_\omega(\cdot) = (1 - \varepsilon)\phi_d(\cdot) + \varepsilon \phi_d\left(\cdot - \frac{r}{\sqrt{d}}\omega\right) \in \mathcal{F}[\rho].$$

Then, we have to propose an upper bound for the term $\mathbb{E}_0[L_\pi^2(\underline{X})]$ where in this setting

$$\begin{aligned} L_\pi(\underline{X}) &= \frac{d\mathbb{P}_\pi}{d\mathbb{P}_0}(\underline{X}) \\ &= \frac{1}{2^d} \sum_{\omega \in \{-1,1\}^d} \prod_{i=1}^n \left[(1 - \varepsilon) + \varepsilon \frac{\phi_d(X_i - \frac{r}{\sqrt{d}}\omega)}{\phi_d(X_i)} \right] \\ &= \frac{1}{2^d} \sum_{\omega \in \{-1,1\}^d} \prod_{i=1}^n \big[(1 - \varepsilon) + \varepsilon e^{-\frac{r^2}{2}} e^{\langle X_i, \frac{r}{\sqrt{d}}\omega \rangle} \big]. \end{aligned}$$

Thus,

$$L_\pi^2(\underline{X}) = \frac{1}{2^{2d}} \sum_{\omega,\tilde{\omega}\in\{-1,1\}^d} \prod_{i=1}^{n} \Big[(1-\varepsilon)^2 + \varepsilon(1-\varepsilon)e^{-\frac{r^2}{2}}\big(e^{\langle X_i, \frac{r}{\sqrt{d}}\omega\rangle} + e^{\langle X_i, \frac{r}{\sqrt{d}}\tilde{\omega}\rangle}\big) + \varepsilon^2 e^{-r^2} e^{\langle X_i, \frac{r}{\sqrt{d}}(\omega+\tilde{\omega})\rangle}\Big].$$

Since for all $\mu \in \mathbb{R}^d$, $\mathbb{E}_0[e^{\langle X_i,\mu\rangle}] = e^{\|\mu\|^2/2}$, we have

$$\mathbb{E}_0\big[L_\pi^2(\underline{X})\big] = \frac{1}{2^{2d}} \sum_{\omega,\tilde{\omega}\in\{-1,1\}^d} \prod_{i=1}^{n} \big[(1-\varepsilon)^2 + 2\varepsilon(1-\varepsilon) + \varepsilon^2 e^{-r^2} e^{\frac{r^2}{2d}\|\omega+\tilde{\omega}\|^2}\big].$$

Noticing that $\|\omega+\tilde{\omega}\|^2 = 2d + 2\langle\omega,\tilde{\omega}\rangle$,

$$\begin{aligned}\mathbb{E}_0\big[L_\pi^2(\underline{X})\big] &= \frac{1}{2^{2d}} \sum_{\omega,\tilde{\omega}\in\{-1,1\}^d} \prod_{i=1}^{n} \big[1 - \varepsilon^2 + \varepsilon^2 e^{\frac{r^2}{d}\langle\omega,\tilde{\omega}\rangle}\big] \\ &= \frac{1}{2^{2d}} \sum_{\omega,\tilde{\omega}\in\{-1,1\}^d} \big[1 + \varepsilon^2\big(e^{\frac{r^2}{d}\langle\omega,\tilde{\omega}\rangle} - 1\big)\big]^n \\ &= \mathbb{E}\big[\big\{1 + \varepsilon^2\big(e^{\frac{r^2}{d}\langle W,\tilde{W}\rangle} - 1\big)\big\}^n\big],\end{aligned}$$

where W and $\tilde{W}$ are two independent d-dimensional Rademacher random variables, i.e.

$$\mathbb{P}(W = w) = \mathbb{P}(\tilde{W} = w) = \frac{1}{2^d} \quad \forall w \in \{-1,1\}^d.$$

Noticing that

$$\langle W, \tilde{W}\rangle = \sum_{j=1}^{d} W_j \tilde{W}_j$$

and that the variables $W_j\tilde{W}_j$ for $1 \le j \le d$ are also i.i.d. Rademacher random variables, $\langle W, \tilde{W}\rangle$ has the same distribution as $Y = \sum_{j=1}^{d} W_j$. This leads to

$$\mathbb{E}_0\big[L_\pi^2(\underline{X})\big] = \mathbb{E}\big[\big\{1 + \varepsilon^2\big(e^{\frac{r^2}{d}Y} - 1\big)\big\}^n\big].$$

Let $C > 0$. We now use the following inequality which holds for any real number u such that $|u| \le C$:

$$\big|e^u - 1 - u\big| \le \frac{e^C}{2}u^2. \tag{13}$$

We set $M = 0.4$, since $\rho \le \rho^{\#}$ and $\varepsilon \ge d^{1/4}/\sqrt{n}$, we have

$$\left|r^2\frac{Y}{d}\right| \le r^2 \le \frac{(\rho^{\#})^2}{\varepsilon^2} \le M^2.$$

Hence, we have

$$e^{\frac{r^2}{d}Y} - 1 \le \frac{r^2}{d}Y + \frac{e^{M^2}}{2}\frac{r^4}{d^2}Y^2.$$

This gives

$$0 \le 1-\varepsilon^2 \le 1+\varepsilon^2\big(e^{\frac{r^2}{d}Y}-1\big) \le 1+\frac{\varepsilon^2 r^2}{\sqrt{d}}\bigg(\frac{Y}{\sqrt{d}}+\frac{e^{M^2}}{2}M^2\frac{Y^2}{d}\bigg).$$

Setting $C(M)=1+e^{M^2}M^2/2$,

$$0 \le 1+\varepsilon^2\big(e^{\frac{r^2}{d}Y}-1\big) \le 1+C(M)\frac{\varepsilon^2 r^2}{\sqrt{d}}\bigg(\frac{|Y|}{\sqrt{d}}\vee\frac{Y^2}{d}\bigg),$$

we obtain

$$\mathbb{E}_0\big[L_\pi^2(\underline{X})\big] \le \mathbb{E}\bigg[\bigg\{1+a\bigg(\frac{|Y|}{\sqrt{d}}\vee\frac{Y^2}{d}\bigg)\bigg\}^n\bigg],$$

where

$$a = C(M)\varepsilon^2 r^2/\sqrt{d}. \tag{14}$$

Using the inequality $\ln(1+x)\le x$ for all $x\ge 0$, we have

$$\begin{aligned}\mathbb{E}_0\big[L_\pi^2(\underline{X})\big] &\le \mathbb{E}\big[e^{\{na(\frac{|Y|}{\sqrt{d}}\vee\frac{Y^2}{d})\}}\big]\\ &\le e^{na}\mathbb{P}\bigg(\frac{|Y|}{\sqrt{d}}\le 1\bigg)+\mathbb{E}\big[e^{na\frac{Y^2}{d}}\mathbb{1}_{\{\frac{|Y|}{\sqrt{d}}>1\}}\big].\end{aligned}$$

Moreover, using an integration by part

$$\mathbb{E}\big[e^{na\frac{Y^2}{d}}\mathbb{1}_{\{\frac{|Y|}{\sqrt{d}}>1\}}\big] \le e^{na}\mathbb{P}\bigg(\frac{|Y|}{\sqrt{d}}>1\bigg)+\int_{e^{na}}^{+\infty}\mathbb{P}\big(e^{na\frac{Y^2}{d}}>t\big)\,dt,$$

leading to

$$\mathbb{E}_0\big[L_\pi^2(\underline{X})\big] \le e^{na}+\int_{e^{na}}^{+\infty}\mathbb{P}\big(e^{na\frac{Y^2}{d}}>t\big)\,dt.$$

We deduce from Hoeffding's inequality that for all $x>0$,

$$\mathbb{P}\bigg(\frac{|Y|}{\sqrt{d}}>x\bigg)\le 2\exp\big(-x^2/2\big).$$

Hence, for all $t>e^{na}$,

$$\mathbb{P}\big(e^{na\frac{Y^2}{d}}>t\big)\le 2t^{-1/2na}.$$

In the particular case where $na<1/2$, we get

$$\begin{aligned}\mathbb{E}_0\big[L_\pi^2(\underline{X})\big] &\le e^{na}+2\int_{e^{na}}^{+\infty}t^{-1/2na}\,dt\\ &\le e^{na}\bigg(1+\frac{4na}{1-2na}e^{-1/2}\bigg)\le h(na),\end{aligned}$$

where the function $h(\cdot)$ is defined as

$$h(x)=e^x\bigg(1+\frac{4x}{1-2x}e^{-1/2}\bigg)\quad\forall x\in[0,1/2[.$$

The function h is non decreasing on $[0, 1/2[$. Hence

$$na \le 1/4 \quad \Rightarrow \quad \mathbb{E}_0\big[L_\pi^2(\underline{X})\big] \le h(1/4) \le 3 < 1 + \eta(\alpha, \beta)^2,$$

since, according to our assumption

$$\alpha + \beta < 1 - \frac{1}{\sqrt{2}} \simeq 0.293 \quad \Rightarrow \quad (1 - \alpha - \beta)^2 > 1/2.$$

In order to conclude the proof, just remark from (14) that

$$na \le 1/4 \quad \Leftrightarrow \quad \rho^2 = \varepsilon^2 r^2 \le \sqrt{d}/\big(4C(M)n\big).$$

Hence, setting $\rho^\# = Md^{1/4}/\sqrt{n}$, the last inequality holds if $M^2 \le 1/(4C(M))$ which is true for $M \le 0.4$.

We finally get that if $\rho \le \rho^\#$, then $\mathbb{E}_0[L_\pi^2(\underline{X})] < 1 + \eta(\alpha, \beta)^2$, which leads to the desired result.

5.2. *Proof of Theorem 6*

In this context,

$$\mathcal{F}[\rho] = \mathcal{F}_\infty[\rho] = \big\{ f \in \mathcal{F};\, \varepsilon \|\mu\|_\infty \ge \rho \big\},$$

for any $\rho > 0$. Let $r > 0$ and $\varepsilon \in]0, 1[$ such that $\varepsilon r = \rho$. In this context, we choose

$$\Omega = \left\{ \omega \in \{0, 1\}^d \text{ s.t. } \sum_{j=1}^d \omega_j = 1 \right\},$$

and we define for all $\omega \in \Omega$,

$$g_\omega(\cdot) = (1 - \varepsilon)\phi_d(\cdot) + \varepsilon\phi_d(\cdot - r\omega) \in \mathcal{F}_\infty[\rho].$$

Now, we turn our attention to the control of the associated likelihood ratio. For each $j = 1, \ldots, d$, let $D^{(j)} \in \{0, 1\}^d$ such that $D_\ell^{(j)} = \mathbb{1}_{\ell=j}$ and

$$\begin{aligned} L_\pi(\underline{X}) &= \frac{d\mathbb{P}_\pi}{d\mathbb{P}_0}(\underline{X}) \\ &= \left[\prod_{i=1}^n \phi_d(X_i)\right]^{-1} \left[\frac{1}{d}\sum_{j=1}^d \prod_{i=1}^n \big\{(1 - \varepsilon)\phi_d(X_i) + \varepsilon\phi_d\big(X_i - rD^{(j)}\big)\big\}\right] \\ &= \frac{1}{d}\sum_{j=1}^d U_j(\underline{X}), \end{aligned}$$

with

$$U_j(\underline{X}) = \prod_{i=1}^n \left\{(1 - \varepsilon) + \varepsilon\frac{\phi_d(X_i - rD^{(j)})}{\phi_d(X_i)}\right\}.$$

Thus

$$\mathbb{E}_0\big[L_\pi^2(\underline{X})\big] = \frac{1}{d^2}\sum_{j=1}^d \mathbb{E}_0\big[U_j(\underline{X})^2\big] + \frac{1}{d^2}\sum_{k \ne j} \mathbb{E}_0\big[U_j(\underline{X})U_k(\underline{X})\big]. \tag{15}$$

In a first time, we can remark that for all $j \in \{1, \dots, d\}$

$$\mathbb{E}_0\big[U_j(\underline{X})^2\big] = \mathbb{E}_0\left[\left(\prod_{i=1}^{n}\left\{(1-\varepsilon) + \varepsilon\frac{\phi_d(X_i - rD^{(j)})}{\phi_d(X_i)}\right\}\right)^2\right]$$
$$= \mathbb{E}_{\phi_d}\left[\left\{(1-\varepsilon) + \varepsilon\frac{\phi_d(X_1 - rD^{(j)})}{\phi_d(X_1)}\right\}^2\right]^n,$$

and

$$\mathbb{E}_{\phi_d}\left[\left\{(1-\varepsilon) + \varepsilon\frac{\phi_d(X - rD^{(j)})}{\phi_d(X)}\right\}^2\right]$$
$$= (1-\varepsilon)^2 + \varepsilon^2\int_{\mathbb{R}^d}\frac{\phi_d^2(x - rD^{(j)})}{\phi_d(x)}\,dx + 2(1-\varepsilon)\varepsilon\int_{\mathbb{R}^d}\phi_d\big(x - rD^{(j)}\big)\,dx$$
$$= (1-\varepsilon)^2 + \varepsilon^2 e^{r^2} + 2(1-\varepsilon)\varepsilon$$
$$= 1 + \varepsilon^2\big(e^{r^2} - 1\big),$$

since $\int_{\mathbb{R}^d}\frac{\phi_d^2(x-\mu)}{\phi_d(x)}\,dx = \exp(\|\mu\|^2)$. Thus

$$\mathbb{E}_0\big[U_j(\underline{X})^2\big] = \big\{1 + \varepsilon^2\big(e^{r^2} - 1\big)\big\}^n.$$

Concerning the second sum in (15), we obtain for all $j, k \in \{1, \dots, d\}$, $j \neq k$

$$\mathbb{E}_0\big[U_j(\underline{X})U_k(\underline{X})\big]$$
$$= \mathbb{E}_0\left[\prod_{i=1}^{n}\left\{(1-\varepsilon) + \varepsilon\frac{\phi_d(X_i - rD^{(j)})}{\phi_d(X_i)}\right\}\left\{(1-\varepsilon) + \varepsilon\frac{\phi_d(X_i - rD^{(k)})}{\phi_d(X_i)}\right\}\right]$$
$$= \left\{\mathbb{E}_{\phi_d}\left[(1-\varepsilon)^2 + (1-\varepsilon)\varepsilon\frac{\phi_d(X_1 - rD^{(j)}) + \phi_d(X_1 - rD^{(k)})}{\phi_d(X_1)}\right.\right.$$
$$\left.\left. + \varepsilon^2\frac{\phi_d(X_1 - rD^{(j)})\phi_d(X_1 - rD^{(k)})}{\phi_d(X_1)^2}\right]\right\}^n$$
$$= \big\{(1-\varepsilon)^2 + 2(1-\varepsilon)\varepsilon + \varepsilon^2\exp\big[r^2\big\langle D^{(j)}, D^{(k)}\big\rangle\big]\big\}^n$$
$$= \big\{(1-\varepsilon)^2 + 2(1-\varepsilon)\varepsilon + \varepsilon^2\big\}^n = 1,$$

since $\int_{\mathbb{R}^d}\frac{\phi_d(x-\mu_1)\phi_d(x-\mu_2)}{\phi_d(x)}\,dx = \exp(\langle\mu_1, \mu_2\rangle)$. Finally,

$$\mathbb{E}_0\big[L_\pi^2(\underline{X})\big] = \frac{1}{d}\big\{1 + \varepsilon^2\big(e^{r^2} - 1\big)\big\}^n + \frac{d(d-1)}{d^2}.$$

We obtain

$$\mathbb{E}_0\big[L_\pi^2(\underline{X})\big] < 1 + \eta(\alpha,\beta)^2 \quad \Leftrightarrow \quad \frac{1}{d}\big\{1 + \varepsilon^2\big(e^{r^2} - 1\big)\big\}^n + \frac{d(d-1)}{d^2} < 1 + \eta(\alpha,\beta)^2$$
$$\Leftrightarrow \quad \big\{1 + \varepsilon^2\big(e^{r^2} - 1\big)\big\}^n < 1 + d\eta(\alpha,\beta)^2. \tag{16}$$

At this step, set

$$r^2 = \ln(1+u) \quad \text{where } u = \frac{1}{n\varepsilon^2}\ln\big(1 + d\eta(\alpha,\beta)^2\big). \tag{17}$$

Then,

$$\begin{aligned}\left\{1+\varepsilon^2\left(e^{r^2}-1\right)\right\}^n &= \left(1+\varepsilon^2 u\right)^n \\ &= \left\{1+\frac{\ln(1+d\eta(\alpha,\beta)^2)}{n}\right\}^n \\ &= \exp\left\{n\ln\left(1+\frac{\ln(1+d\eta(\alpha,\beta)^2)}{n}\right)\right\} \\ &\leq 1+d\eta(\alpha,\beta)^2,\end{aligned}$$

where for the last line, we have used the inequality $\ln(1+x)\leq x$ for all $x\geq 0$. Hence, Inequality (16) is satisfied provided r is chosen according to (17). This concludes the proof of Theorem 6.

6. Proof of the upper bounds

6.1. *Proof of Theorem 2*

First, remark that

$$\|\sqrt{n}\bar{X}_n\|^2 = \sum_{j=1}^{d}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}X_{ij}\right)^2.$$

Under H_0, X_{ij} are i.i.d. standard Gaussian random variables. Hence $\|\sqrt{n}\bar{X}_n\|^2$ is a chi-square random variable with d degrees of freedom and

$$\mathbb{P}_0(\Psi_{1,\alpha}=1)=\mathbb{P}_0\left(\|\sqrt{n}\bar{X}_n\|^2>t_{d,\alpha}\right)=\alpha,$$

according to the definition of the quantile $t_{d,\alpha}$. The test $\Psi_{1,\alpha}$ is hence of level α.

Now, we want to control the second kind error. Under H_1, each variable X_i can be written as

$$X_i = V_i\mu+\eta_i,$$

where V_i is a Bernoulli variable with parameter ε, $\eta_i\sim\mathcal{N}_d(0_d,I_d)$ and V_i and η_i are independent. Then

$$\sqrt{n}\bar{X}_n=\frac{S}{\sqrt{n}}\mu+B,$$

where $S=\sum_{i=1}^{n}V_i\sim\mathcal{B}(n,\varepsilon)$ is a binomial random variable with parameters (n,ε), $B=\sum_{i=1}^{n}\eta_i/\sqrt{n}\sim\mathcal{N}_d(0_d,I_d)$ and S, B are independent. In particular, conditionally to S, the variable $\|\sqrt{n}\bar{X}_n\|^2=\|\frac{S}{\sqrt{n}}\mu+B\|^2$ has a non-central chi-square distribution with d degrees of freedom and noncentrality parameter $\lambda_S=\|\frac{S}{\sqrt{n}}\mu\|^2$. Introduce

$$h_S=d+\lambda_S-2\sqrt{[d+2\lambda_S]\ln(2/\beta)}.$$

According to Lemma 2 in Appendix A [see also [17]]

$$\mathbb{P}\left(\left\|\frac{S}{\sqrt{n}}\mu+B\right\|^2\leq h_S|S\right)\leq\frac{\beta}{2}.$$

Hence, for each $f \in \mathcal{F}$,

$$\begin{aligned}
\mathbb{P}_f(\Psi_{1,\alpha} = 0) &= \mathbb{P}_f\big(\|\sqrt{n}\bar{X}_n\|^2 \le t_{d,\alpha}\big) \\
&= \mathbb{P}\left(\left\|\frac{S}{\sqrt{n}}\mu + B\right\|^2 \le t_{d,\alpha}\right) \\
&= \mathbb{P}\left(\left\{\left\|\frac{S}{\sqrt{n}}\mu + B\right\|^2 \le t_{d,\alpha}\right\} \cap \{h_S \le t_{d,\alpha}\}\right) + \mathbb{P}\left(\left\{\left\|\frac{S}{\sqrt{n}}\mu + B\right\|^2 \le t_{d,\alpha}\right\} \cap \{h_S > t_{d,\alpha}\}\right) \\
&\le \mathbb{P}(h_S \le t_{d,\alpha}) + \mathbb{P}\left(\left\|\frac{S}{\sqrt{n}}\mu + B\right\|^2 \le h_S\right) \\
&\le \mathbb{P}(h_S \le t_{d,\alpha}) + \frac{\beta}{2}.
\end{aligned}$$

According to Lemma 1 in Appendix A,

$$t_{d,\alpha} \le d + b(\alpha, d) \quad \text{where } b(\alpha, d) = 2\ln(1/\alpha) + 2\sqrt{d\ln(1/\alpha)}.$$

Hence

$$\begin{aligned}
\mathbb{P}(h_S \le t_{d,\alpha}) &\le \mathbb{P}\big(h_S \le d + b(d,\alpha)\big) \\
&\le \mathbb{P}\big(\lambda_S - 2\sqrt{[d + 2\lambda_S]\ln(2/\beta)} \le b(d,\alpha)\big) \\
&\le \mathbb{P}\big(\lambda_S - 2\sqrt{2\ln(2/\beta)}\sqrt{\lambda_S} - \big[2\sqrt{d\ln(2/\beta)} + b(d,\alpha)\big] \le 0\big) \\
&\le \mathbb{P}\big(\sqrt{\lambda_S} \le R(\alpha,\beta,d)\big),
\end{aligned}$$

with

$$R(\alpha,\beta,d) = \sqrt{2\ln(2/\beta)} + \sqrt{2\ln(2/\beta) + 2\sqrt{d\ln(2/\beta)} + b(\alpha,d)}.$$

We notice that $R(\alpha,\beta,d) \le C(\alpha,\beta)d^{1/4}$ where $C(\alpha,\beta)$ is a constant only depending on α and β. Assuming that $\sqrt{n}\varepsilon\|\mu\| > C(\alpha,\beta)d^{1/4}$ and using a Tchebychev's inequality leads to

$$\begin{aligned}
\mathbb{P}(h_S \le t_{d,\alpha}) &\le \mathbb{P}\left(S \le \frac{\sqrt{n}}{\|\mu\|}C(\alpha,\beta)d^{1/4}\right) \\
&\le \mathbb{P}\left(|S - n\varepsilon| > n\varepsilon - \frac{\sqrt{n}}{\|\mu\|}C(\alpha,\beta)d^{1/4}\right) \\
&\le \frac{n\varepsilon\|\mu\|^2}{[n\varepsilon\|\mu\| - \sqrt{n}C(\alpha,\beta)d^{1/4}]^2}. \qquad (18)
\end{aligned}$$

If $\sqrt{n}\varepsilon\|\mu\| > 2C(\alpha,\beta)d^{\frac{1}{4}}$ then $n\varepsilon\|\mu\| - \sqrt{n}C(\alpha,\beta)d^{1/4} \ge \frac{n\varepsilon\|\mu\|}{2}$.
Thus, $\mathbb{P}(h_S \le t_{d,\alpha}) \le \frac{4}{n\varepsilon} \le \frac{\beta}{2}$ as soon as $n\varepsilon \ge \frac{8}{\beta}$.

6.2. *Proof of Theorem 3*

It is easy to see that $\Psi_{2,\alpha}$ is a level-α test. Indeed,

$$\begin{aligned}
\mathbb{P}_0(\Psi_{2,\alpha} = 1) &= \mathbb{P}_0\big(\exists 1 \le i \le n, \|X_i\|^2 > t_{d,\alpha/n}\big) \\
&\le \sum_{i=1}^{n} \mathbb{P}_0\big(\|X_i\|^2 > t_{d,\alpha/n}\big) \le \sum_{i=1}^{n} \frac{\alpha}{n} = \alpha
\end{aligned}$$

since, under the null hypothesis, $\|X_i\|^2$ is a chi-square variable with d degrees of freedom. Note that, according to Lemma 1 in Appendix A,

$$t_{d,\alpha/n} \le d + 2\sqrt{d\ln(n/\alpha)} + 2\ln(n/\alpha).$$

Then, under H_1, each variable X_i can be written as

$$X_i = V_i\mu + \eta_i, \quad i = 1, \dots, n,$$

where $V_i \sim \mathcal{B}(\varepsilon)$ denotes a random Bernoulli variable and $\eta_i \sim \mathcal{N}_d(0_d, I_d)$, V_i and η_i being independent. Let $S = \sum_{i=1}^n V_i \sim \mathcal{B}(n, \varepsilon)$. We have

$$\mathbb{P}_f(\Psi_{2,\alpha} = 0) \le \mathbb{P}_f(\Psi_{2,\alpha} = 0 \cap S \ge n\varepsilon/2) + \mathbb{P}(S \le n\varepsilon/2).$$

First, according to Markov's inequality,

$$\mathbb{P}\left(|S - n\varepsilon| \ge \sqrt{\frac{2n\varepsilon}{\beta}}\right) \le \frac{\beta}{2},$$

thus $\mathbb{P}(S < n\varepsilon - \sqrt{\frac{2n\varepsilon}{\beta}}) \le \frac{\beta}{2}$. Assuming that $n\varepsilon \ge \frac{8}{\beta}$,

$$\mathbb{P}\left(S \le \frac{n\varepsilon}{2}\right) \le \mathbb{P}\left(S < n\varepsilon - \sqrt{\frac{2n\varepsilon}{\beta}}\right) \le \frac{\beta}{2}.$$

Second, since $n\varepsilon/2 \ge \frac{4}{\beta} \ge 1$,

$$\begin{aligned}
\mathbb{P}_f(\Psi_{2,\alpha} = 0 \cap S \ge n\varepsilon/2) &\le \mathbb{P}_f(\Psi_{2,\alpha} = 0 \cap S \ge 1) \\
&\le \mathbb{P}_f\big(\exists 1 \le i \le n, \|\mu + \eta_i\|^2 \le t_{d,\alpha/n}\big) \\
&\le \sum_{i=1}^n \mathbb{P}_f\big(\|\mu + \eta_i\|^2 \le t_{d,\alpha/n}\big) \\
&\le n\mathbb{P}_f\big(\|\mu + \eta_1\|^2 \le t_{d,\alpha/n}\big),
\end{aligned}$$

where $\|\mu + \eta_1\|^2$ is a noncentral chi-square random variable with d degrees of freedom and a noncentrality parameter $\|\mu\|^2$. We deduce from Lemma 2 that for $x_\beta = \ln(2n/\beta)$,

$$\mathbb{P}\big(\|\mu + \eta_1\|^2 \le d + \|\mu\|^2 - 2\sqrt{\big(d + 2\|\mu\|^2\big)x_\beta}\big) \le e^{-x_\beta} = \frac{\beta}{2n}.$$

Gathering the previous inequalities, we get that $\mathbb{P}_f(\Psi_{2,\alpha} = 0) \le \beta$ provided that

$$t_{d,\alpha/n} \le d + \|\mu\|^2 - 2\sqrt{\big(d + 2\|\mu\|^2\big)x_\beta}.$$

After some easy computations, we see that this condition is fulfilled if

$$\|\mu\|^2 \ge C(\alpha, \beta)\big(\sqrt{d\ln(n)} + \ln(n)\big),$$

which concludes the proof.

6.3. *Proof of Theorem 5*

Following the definition of $t_{n,d,k,\alpha}$ and since T_U has a chi-square distribution with d degrees of freedom under H_0, $\Psi_{3,\alpha}$ is ensured to be a level-α test:

$$\begin{aligned}\mathbb{P}_0(\Psi_{3,\alpha}=1) &\le \sum_{k=1}^{n}\sum_{U;|U|=k}\mathbb{P}_0(T_U > t_{n,d,k,\alpha})\\ &\le \sum_{k=1}^{n}\sum_{U;|U|=k} e^{-x_{n,k,\alpha}}\\ &\le \sum_{k=1}^{n}\binom{n}{k}\left(\frac{en}{k}\right)^{-k}\frac{\alpha}{n}\le\alpha,\end{aligned}$$

according to Lemma 1.

Now, we want to control the second kind error. Let $f\in\mathcal{F}$. Under H_1, each variable X_i can be written as $X_i = V_i\mu+\eta_i$ where V_i is a Bernoulli variable with parameter ε, independent of $\eta_i\sim\mathcal{N}_d(0_d, I_d)$. Let $S=\sum_{i=1}^{n}V_i\sim\mathcal{B}(n,\varepsilon)$ and for all $U\in\mathcal{U}$, $S_U=\sum_{i\in U}V_i\sim\mathcal{B}(|U|,\varepsilon)$. The second kind error can be upper bounded by

$$\mathbb{P}_f(\Psi_{3,\alpha}=0)\le\mathbb{P}_f(\Psi_{3,\alpha}=0\cap S\ge n\varepsilon/2)+\mathbb{P}(S\le n\varepsilon/2). \tag{19}$$

First, according to Markov's inequality, we get, as in the proof of Theorem 3 that

$$\mathbb{P}\left(S\le\frac{n\varepsilon}{2}\right)\le\frac{\beta}{2},$$

since $n\varepsilon\ge\frac{8}{\beta}$. Second,

$$\mathbb{P}_f(\Psi_{3,\alpha}=0\cap S\ge n\varepsilon/2)=\mathbb{P}_f\big(\forall k\in\{1,\dots,n\},\forall U;|U|=kT_U\le t_{n,d,k,\alpha}\cap S\ge n\varepsilon/2\big).$$

Let $\tilde{k}=\lceil n\varepsilon\rceil$. If $S\ge\frac{n\varepsilon}{2}$, there exists $U_0\in\mathcal{U}$ such that $|U_0|=\tilde{k}$ and $S_{U_0}\ge\frac{\tilde{k}}{2}$. Since

$$T_{U_0}=\left\|\frac{1}{\sqrt{|U_0|}}\sum_{i\in U_0}\eta_i+\frac{S_{U_0}}{\sqrt{|U_0|}}\mu\right\|^2,$$

$T_{U_0}|S_{U_0}\sim\chi^2(d,\frac{S_{U_0}^2}{\tilde{k}}\|\mu\|^2)$. Thus,

$$\begin{aligned}\mathbb{P}_f(\Psi_{3,\alpha}=0\cap S\ge n\varepsilon/2)&\le\mathbb{P}_f\left(\exists U_0;|U_0|=\tilde{k};T_{U_0}\le t_{n,d,\tilde{k},\alpha}\cap S_{U_0}\ge\frac{\tilde{k}}{2}\right)\\ &\le\sum_{U_0;|U_0|=\tilde{k}}\mathbb{P}_f\left(T_{U_0}\le t_{n,d,\tilde{k},\alpha}\cap S_{U_0}\ge\frac{\tilde{k}}{2}\right).\end{aligned}$$

We remark that $\frac{S_{U_0}^2}{\tilde{k}}\|\mu\|^2\ge\frac{\tilde{k}}{4}\|\mu\|^2$ if $S_{U_0}\ge\frac{\tilde{k}}{2}$. Thus, according to Lemma 3,

$$\mathbb{P}_f\left(T_{U_0}\le t_{n,d,\tilde{k},\alpha}\cap S_{U_0}\ge\frac{\tilde{k}}{2}\right)\le\mathbb{P}(A\le t_{n,d,\tilde{k},\alpha}),$$

where $A \sim \chi^2(d, \frac{\tilde{k}}{4}\|\mu\|^2)$. Then,

$$\begin{aligned}
\mathbb{P}_f(\Psi_{3,\alpha} = 0 \cap S \geq n\varepsilon/2) &\leq \sum_{U_0;|U_0|=\tilde{k}} \mathbb{P}_f\left(T_{U_0} \leq t_{n,d,\tilde{k},\alpha} \cap S_{U_0} \geq \frac{\tilde{k}}{2}\right) \\
&\leq \binom{n}{\tilde{k}} \mathbb{P}(A \leq t_{n,d,\tilde{k},\alpha}) \\
&\leq \left(\frac{en}{\tilde{k}}\right)^{\tilde{k}} \mathbb{P}(A \leq \boldsymbol{\kappa}) \quad \text{if } t_{n,d,\tilde{k},\alpha} \leq \boldsymbol{\kappa} \\
&\leq \left(\frac{en}{\tilde{k}}\right)^{\tilde{k}} e^{-\tilde{x}} = \frac{\beta}{2},
\end{aligned}$$

where $\boldsymbol{\kappa} = d + \frac{\tilde{k}}{4}\|\mu\|^2 - 2\sqrt{(d + \frac{\tilde{k}}{4}\|\mu\|^2)\tilde{x}}$ and $\tilde{x} = \tilde{k}\ln(en/\tilde{k}) + \ln(\beta/2)$. The condition $t_{n,d,\tilde{k},\alpha} \leq \boldsymbol{\kappa}$ is equivalent to

$$\begin{aligned}
t_{n,d,\tilde{k},\alpha} \leq \boldsymbol{\kappa} \quad &\Leftrightarrow \quad d + 2\sqrt{dx_{n,\tilde{k},\alpha}} + 2x_{n,\tilde{k},\alpha} \leq d + \frac{\tilde{k}}{4}\|\mu\|^2 - 2\sqrt{\left(d + \frac{\tilde{k}}{4}\|\mu\|^2\right)\tilde{x}} \\
&\Leftrightarrow \quad \sqrt{dx_{n,\tilde{k},\alpha}} + x_{n,\tilde{k},\alpha} + \sqrt{\left(d + \frac{\tilde{k}}{4}\|\mu\|^2\right)\tilde{x}} \leq \frac{\tilde{k}}{8}\|\mu\|^2. \qquad (20)
\end{aligned}$$

After some easy computations, one can show that Condition (20) is satisfied if

$$\|\mu\|^2 \geq \frac{16}{\tilde{k}}(\tilde{x} + \sqrt{d\tilde{x}} + x_{n,\tilde{k},\alpha} + \sqrt{dx_{n,\tilde{k},\alpha}}).$$

Noting that $x_{n,\tilde{k},\alpha} \leq c(\alpha)n\varepsilon\ln(1/\varepsilon)$ and $\tilde{x} \leq c(\beta)n\varepsilon\ln(1/\varepsilon)$, and since $\tilde{k} = [n\varepsilon]$, (20) holds as soon as

$$\|\mu\|^2 \geq C(\alpha,\beta)\left[\ln\left(\frac{1}{\varepsilon}\right) + \sqrt{\frac{d}{n\varepsilon}\ln\left(\frac{1}{\varepsilon}\right)}\right].$$

We now consider the aggregated test

$$\Psi_\alpha = \Psi_{1,\alpha/2} \vee \Psi_{3,\alpha/2}.$$

Then, $\mathbb{P}_f(\Psi_\alpha = 0) \leq \beta$ if

$$\varepsilon^2\|\mu\|^2 \geq C(\alpha,\beta)\left[\left(\frac{\sqrt{d}}{n}\right) \wedge \left\{\varepsilon^2\ln\left(\frac{1}{\varepsilon}\right) + \varepsilon^{3/2}\sqrt{\frac{d}{n}\ln\left(\frac{1}{\varepsilon}\right)}\right\}\right].$$

Let us now compute the right hand term of the previous inequality.

- If $n\varepsilon^3\ln(1/\varepsilon) \geq 1$, we get

$$C(\alpha,\beta)\left(\frac{\sqrt{d}}{n}\right).$$

- If $n^{-1/2} \leq \varepsilon$ and $n\varepsilon^3\ln(1/\varepsilon) < 1$, then
  - If $d \leq n^2\varepsilon^4\ln^2(\frac{1}{\varepsilon})$ we get

$$C(\alpha,\beta)\left(\frac{\sqrt{d}}{n}\right).$$

– If $n^2\varepsilon^4 \ln^2(\frac{1}{\varepsilon}) \le d \le n\varepsilon \ln(\frac{1}{\varepsilon})$, we get

$$C(\alpha,\beta)\varepsilon^2 \ln\left(\frac{1}{\varepsilon}\right).$$

– If $d \ge n\varepsilon \ln(\frac{1}{\varepsilon})$ we obtain

$$C(\alpha,\beta)\varepsilon^{3/2}\sqrt{\frac{d}{n}\ln\left(\frac{1}{\varepsilon}\right)}.$$

6.4. *Proof of Theorem 7*

First, $\Psi_{4,\alpha}$ is a level-α test since

$$\begin{aligned}\mathbb{P}_0(\Psi_{4,\alpha}=1) &= \mathbb{P}_0\big(\exists j \in \{1,\dots,d\}; n\bar{X}_j^2 > t_{1,\frac{\alpha}{d}}\big)\\ &\le \sum_{j=1}^{d} \mathbb{P}_0\big(n\bar{X}_j^2 > t_{1,\frac{\alpha}{d}}\big) = d \times \frac{\alpha}{d} = \alpha\end{aligned}$$

since $n\bar{X}_j^2$ is a chi-square variable with one degree of freedom under H_0.

Second, let $f = f_{(\varepsilon,\mu)} \in \mathcal{F}$. According to Lemma 4 and since $n\varepsilon \ge \frac{8}{\beta}$, the second kind error is controlled by

$$\begin{aligned}\mathbb{P}_f(\Psi_{4,\alpha}=0) &= \mathbb{P}_f\big(\forall j \in \{1,\dots,d\}; n\bar{X}_j^2 \le t_{1,\alpha/d}\big)\\ &\le \inf_j \mathbb{P}_f\big(n\bar{X}_j^2 \le t_{1,\alpha/d}\big)\\ &\le \beta\end{aligned}$$

if

$$\exists j \in \{1,\dots,d\}; \quad \sqrt{n}\varepsilon|\mu_j| > C_\beta + C\sqrt{\ln(d/\alpha)}.$$

Thus, $\mathbb{P}_f(\Psi_{4,\alpha}=0) \le \beta$ if

$$\varepsilon\|\mu\|_\infty \ge C(\alpha,\beta)\sqrt{\frac{\ln(d)}{n}},$$

where $C(\alpha,\beta)$ is a positive constant only depending on α and β.

6.5. *Proof of Theorem 8*

First, $\Psi_{5,\alpha}$ is a level-α test since

$$\begin{aligned}\mathbb{P}_0(\Psi_{5,\alpha}=1) &= \mathbb{P}_0\big(\exists j \in \{1,\dots,d\}, \exists i \in \{1,\dots,n\}; X_{ij}^2 > t_{1,\frac{\alpha}{nd}}\big)\\ &\le \sum_{j=1}^{d}\sum_{i=1}^{n} \mathbb{P}_0\big(X_{ij}^2 > t_{1,\frac{\alpha}{nd}}\big) = dn \times \frac{\alpha}{dn} = \alpha\end{aligned}$$

since for all i, j, X_{ij}^2 is a chi-square variable with one degree of freedom under H_0.

In a second time, let $f = f_{(\varepsilon,\mu)} \in \mathcal{F}$. According to Lemma 5 and since $n\varepsilon \geq \frac{8}{\beta}$, the second kind error is controlled by

$$\begin{aligned}\mathbb{P}_f(\Psi_{5,\alpha} = 0) &= \mathbb{P}_f\Big(\forall j \in \{1,\dots,d\}; \max_{i=1,\dots,n} X_{ij}^2 \leq t_{1,\alpha/nd}\Big)\\ &\leq \inf_j \mathbb{P}_f\Big(\max_{i=1,\dots,n} X_{ij}^2 \leq t_{1,\alpha/nd}\Big)\\ &\leq \beta\end{aligned}$$

if

$$\exists j \in \{1,\dots,d\} \text{ s.t. } |\mu_j| \geq C_\beta\sqrt{\ln(n)} + \sqrt{C_\beta\sqrt{\ln(n)} + A(\alpha/d,n)},$$

where $A(\frac{\alpha}{d},n) = 2\sqrt{\ln(nd/\alpha)} + 2\ln(nd/\alpha)$. Thus, $\mathbb{P}_f(\Psi_{5,\alpha} = 0) \leq \beta$ if

$$\|\mu\|_\infty \geq C(\alpha,\beta)\sqrt{\ln(dn)},$$

where $C(\alpha,\beta)$ is a positive constant only depending on α and β.

Appendix A: Properties for chi-square distributions and noncentral chi-square distributions

In this section, we present some well-known results there are useful throughout the proofs. The first lemma is concerned with deviation of a chi-square random variable, proposed in [19].

Lemma 1. *Let U be a chi-square random variable with d degrees of freedom. Then,*

- *for any positive x,*

$$\begin{cases}\mathbb{P}(U \geq d + 2\sqrt{dx} + 2x) \leq e^{-x},\\ \mathbb{P}(U \leq d - 2\sqrt{dx}) \leq e^{-x}.\end{cases}$$

- *For any given $\alpha \in]0,1[$, let $u(d,\alpha)$ be the $(1-\alpha)$-quantile of $\chi^2(d)$. Then*

$$u(d,\alpha) \leq d + 2\ln(1/\alpha) + 2\sqrt{d\ln(1/\alpha)} = d + b(\alpha,d).$$

This second lemma provides the control of deviations of a noncentral chi-square random variable, available in [6].

Lemma 2. *Let T be a noncentral chi-square random variable with d degrees of freedom and a noncentrality parameter λ. Then, for any positive x,*

$$\begin{cases}\mathbb{P}(T \geq d + \lambda + 2\sqrt{(d+2\lambda)x} + 2x) \leq e^{-x},\\ \mathbb{P}(T \leq d + \lambda - 2\sqrt{(d+2\lambda)x}) \leq e^{-x}.\end{cases}$$

This third lemma provides that the cumulative distribution function of the noncentral chi-square distribution is a non-increasing function in the noncentrality parameter λ, for a fixed degree of freedom (see for instance [13]).

Lemma 3. *Let T_λ be a noncentral chi-square random variable with d degrees of freedom and a noncentrality parameter λ. Then, if $\lambda \geq \tilde{\lambda}$,*

$$\mathbb{P}(T_\lambda \leq x) \leq \mathbb{P}(T_{\tilde{\lambda}} \leq x)$$

for any real number x.

Appendix B: Unidimensional test

In this section, we consider two testing procedures in a unidimensional context, which are required to prove Theorems 7 and 8.

Let $(Z_1, \dots, Z_n)$ be i.i.d. random variables from an unknown density g w.r.t. the Lebesgue measure on $\mathbb{R}$. We want to test

$$H_0 : g = \phi(\cdot) \quad \text{versus } H_1 : g \in \mathcal{G},$$

where $\phi(\cdot) = \phi_1(\cdot)$ is the unidimensional standard Gaussian density and

$$\mathcal{G} = \left\{ g_{(\varepsilon,\tau)} : z \in \mathbb{R} \mapsto (1-\varepsilon)\phi(z) + \varepsilon\phi(z-\tau); \varepsilon \in]0,1[, \tau \in \mathbb{R} \right\}.$$

B.1. *First testing procedure*

The first procedure is based on the mean $\bar{Z}_n = \frac{1}{n}\sum_{i=1}^n Z_i$. Let $\alpha \in]0,1[$ and $T_{4,\alpha}$ be the testing procedure defined as

$$T_{4,\alpha} = \mathbb{1}_{\{n\bar{Z}_n^2 > t_{1,\alpha}\}}, \tag{21}$$

where $t_{1,\alpha}$ is the $(1-\alpha)$-quantile of a chi-square distribution with one degree of freedom. The following lemma establishes sufficient conditions that allow to control the second kind error of $T_{4,\alpha}$.

Lemma 4. *Let $\alpha, \beta \in]0,1[$. Assume that $n\varepsilon \geq \frac{8}{\beta}$. Then, the testing procedure $T_{4,\alpha}$ defined in (21) is of level α. Moreover, there exist two positive constants C_β and C such that for all $g \in \mathcal{G}$ for which*

$$\sqrt{n}\varepsilon|\tau| > C_\beta + C\sqrt{\ln(1/\alpha)},$$
$$\mathbb{P}_g(T_{4,\alpha} = 0) \leq \beta.$$

Proof. First, it is easy to see that $T_{4,\alpha}$ is a level-α test. Now, we want to control the second kind error. Under H_1, each variable Z_i can be written as

$$Z_i = V_i\tau + \eta_i,$$

where V_i is a Bernoulli variable with parameter ε, $\eta_i \sim \mathcal{N}(0,1)$ and V_i and η_i are independent. Then

$$\sqrt{n}\bar{Z}_n = \frac{S}{\sqrt{n}}\tau + B,$$

where $S = \sum_{i=1}^n V_i \sim \mathcal{B}(n,\varepsilon)$, $B = \sum_{i=1}^n \eta_i/\sqrt{n} \sim \mathcal{N}(0,1)$ and S, B are independent. Conditionally to S, the variable $n\bar{Z}_n^2$ has a non-central chi-square distribution with one degree of freedom and noncentrality parameter $\lambda_S = (\frac{S}{\sqrt{n}}\tau)^2$. Let

$$h_S = 1 + \lambda_S - 2\sqrt{(1+2\lambda_S)\ln(2/\beta)}.$$

According to Lemma 2 in Appendix A (see also [17])

$$\mathbb{P}\left(n\bar{Z}_n^2 \leq h_S | S\right) \leq \frac{\beta}{2}.$$

Hence, for each $g \in \mathcal{G}$,

$$\begin{aligned}\mathbb{P}_g(T_{4,\alpha} = 0) &= \mathbb{P}_g\left(n\bar{Z}_n^2 \leq t_{1,\alpha}\right) \\ &= \mathbb{P}\left(\left\{n\bar{Z}_n^2 \leq t_{1,\alpha}\right\} \cap \{h_S \leq t_{1,\alpha}\}\right)\end{aligned}$$

$$+\mathbb{P}\big(\big\{n\bar{Z}_n^2 \le t_{1,\alpha}\big\}\cap\{h_S > t_{1,\alpha}\}\big)$$
$$\le \mathbb{P}(h_S \le t_{1,\alpha}) + \frac{\beta}{2}.$$

According to Lemma 1 in Appendix A,

$$t_{1,\alpha} \le 1 + 2\ln(1/\alpha) + 2\sqrt{\ln(1/\alpha)}.$$

Hence

$$\mathbb{P}(h_S \le t_{1,\alpha}) \le \mathbb{P}\big(1+\lambda_S - 2\sqrt{(1+2\lambda_S)\ln(2/\beta)} \le 1 + 2\ln(1/\alpha) + 2\sqrt{\ln(1/\alpha)}\big)$$
$$\le \mathbb{P}\big(\sqrt{\lambda_S} \le R(\alpha,\beta,1)\big),$$

with

$$R(\alpha,\beta,1) = \sqrt{2\ln(2/\beta)} + \sqrt{2\ln(2/\beta) + 2\sqrt{\ln(2/\beta)} + 2\ln(1/\alpha) + 2\sqrt{\ln(1/\alpha)}} \le C_\beta + C\sqrt{\ln(1/\alpha)},$$

where C_β and C are two positive constants.

Using a Tchebychev's inequality leads to

$$\mathbb{P}(h_S \le t_{1,\alpha}) \le \mathbb{P}\left(S \le \frac{\sqrt{n}}{|\tau|}\big(C_\beta + C\sqrt{\ln(1/\alpha)}\big)\right)$$
$$\le \mathbb{P}\left(|S - n\varepsilon| > n\varepsilon - \frac{\sqrt{n}}{|\tau|}\big(C_\beta + C\sqrt{\ln(1/\alpha)}\big)\right)$$
$$\le \frac{n\varepsilon\tau^2}{[n\varepsilon|\tau| - \sqrt{n}(C_\beta + C\sqrt{\ln(1/\alpha)})]^2}.$$

If $\sqrt{n}\varepsilon|\tau| > 2(C_\beta + C\sqrt{\ln(1/\alpha)})$ then $n\varepsilon|\tau| - \sqrt{n}(C_\beta + C\sqrt{\ln(1/\alpha)}) \ge \frac{n\varepsilon|\tau|}{2}$. Thus, $\mathbb{P}(h_S \le t_{1,\alpha}) \le \frac{4}{n\varepsilon} \le \frac{\beta}{2}$ as soon as $n\varepsilon \ge \frac{8}{\beta}$. □

B.2. Second testing procedure

Let $\alpha \in]0,1[$ and $T_{5,\alpha}$ be the testing procedure defined as

$$T_{5,\alpha} = \mathbb{1}_{\{\max_{1\le i\le n} Z_i^2 > t_{1,\frac{\alpha}{n}}\}}, \tag{22}$$

where $t_{1,\frac{\alpha}{n}}$ is the $(1-\frac{\alpha}{n})$-quantile of a chi-square distribution with one degree of freedom. The following lemma establishes sufficient conditions that allow to control the second kind error of $T_{5,\alpha}$.

Lemma 5. *Let $\alpha,\beta \in]0,1[$. Assume that $n\varepsilon \ge \frac{8}{\beta}$. Then, the testing procedure $T_{5,\alpha}$ defined in (22) is of level α. Moreover, there exists a positive constant C_β such that for all $g \in \mathcal{G}$ for which*

$$|\tau| \ge C_\beta\sqrt{\ln(n)} + \sqrt{A(\alpha,n)}$$

with $A(\alpha,n) = 2\sqrt{\ln(n/\alpha)} + 2\ln(n/\alpha)$,

$$\mathbb{P}_g(T_{5,\alpha} = 0) \le \beta.$$

Proof. First

$$\begin{aligned}\mathbb{P}_0(T_{5,\alpha}=1) &= \mathbb{P}_0\big(\exists 1\le i\le n, Z_i^2 > t_{1,\alpha/n}\big)\\ &\le \sum_{i=1}^{n}\mathbb{P}_0\big(Z_i^2 > t_{1,\alpha/n}\big)\\ &\le \sum_{i=1}^{n}\frac{\alpha}{n}=\alpha\end{aligned}$$

since, under the null hypothesis, Z_i^2 is a chi-square variable with one degree of freedom. Note that, according to Lemma 1 in Appendix A,

$$t_{1,\alpha/n}\le 1+2\sqrt{\ln(n/\alpha)}+2\ln(n/\alpha).$$

Under H_1, each variable Z_i can be written as

$$Z_i = V_i\tau+\eta_i,\quad i=1,\dots,n,$$

where $V_i\sim\mathcal{B}(\varepsilon)$, $\eta_i\sim\mathcal{N}(0,1)$, and V_i and η_i are independent. Let $S=\sum_{i=1}^{n}V_i\sim\mathcal{B}(n,\varepsilon)$. As in the proof of Theorem 3, assuming that $n\varepsilon\ge\frac{8}{\beta}$, $\mathbb{P}(S\le\frac{n\varepsilon}{2})\le\frac{\beta}{2}$. Moreover,

$$\begin{aligned}\mathbb{P}_g(T_{5,\alpha}=0\cap S\ge n\varepsilon/2) &\le \mathbb{P}_g(T_{5,\alpha}=0\cap S\ge 1)\\ &\le \mathbb{P}\big(\exists 1\le i\le n,(\tau+\eta_i)^2\le t_{1,\alpha/n}\big)\\ &\le n\mathbb{P}\big((\tau+\eta_1)^2\le t_{1,\alpha/n}\big),\end{aligned}$$

where $(\tau+\eta_1)^2$ is a noncentral chi-square random variable with one degree of freedom and a noncentrality parameter τ^2. We deduce from Lemma 2 that for $x_\beta=\ln(2n/\beta)$,

$$\mathbb{P}\big((\tau+\eta_1)^2\le 1+\tau^2-2\sqrt{(1+2\tau^2)x_\beta}\big)\le e^{-x_\beta}=\beta/(2n).$$

Gathering the previous inequalities, we get that $\mathbb{P}_g(T_{5,\alpha}=0)\le\beta$ provided that

$$1+2\sqrt{\ln(n/\alpha)}+2\ln(n/\alpha)\le 1+\tau^2-2\sqrt{(1+2\tau^2)x_\beta}.$$

This condition is fulfilled if

$$\tau^2-C_\beta\sqrt{\ln(n)}|\tau|-A(\alpha,n)\ge 0,$$

where $A(\alpha,n)=2\sqrt{\ln(n/\alpha)}+2\ln(n/\alpha)$ and C_β is a positive constant only depends on β. After some easy computations, we see that this condition is fulfilled if

$$|\tau|\ge C_\beta\sqrt{\ln(n)}+\sqrt{A(\alpha,n)}.$$ □

Acknowledgements

This work was supported by the French Agence Nationale de la Recherche (ANR-13-JS01-0001-01, project MixStatSeq).

We thank Nicolas Verzelen for very helpful discussions that allowed us to improve the paper. We also thank the associate editor and two anonymous referees for their constructive remarks and comments on the first versions of the paper.

References

[1] E. Arias-Castro, E. J. Candes and A. Durand. Detection of an anomalous cluster in a network. *Ann. Statist.* **39** (2011) 278–304. MR2797847

[2] J.-M. Azaïs, É. Gassiat and C. Mercadier. The likelihood ratio test for general mixture models with or without structural parameter. *ESAIM Probab. Stat.* **13** (2009) 301–327. MR2528086

[3] Y. Baraud. Non-asymptotic minimax rates of testing in signal detection. *Bernoulli* **8** (5) (2002) 577–606. MR1935648

[4] Q. Berthet and P. Rigollet. Complexity theoretic lower bounds for sparse principal component detection. *Proceedings of the 26th Annual Conference on Learning Theory* **30** (2013) 1046–1066.

[5] Q. Berthet and P. Rigollet. Optimal detection of sparse principal components in high dimension. *Ann. Statist.* **41** (4) (2013) 1780–1815. MR3127849

[6] L. Birgé. An alternative point of view on Lepski's method. In *State of the Art in Probability and Statistics (Leiden, 1999)* 113–133. *IMS Lecture Notes Monogr. Ser.* **36**. Inst. Math. Statist., Beachwood, OH, 2001. MR1836557

[7] C. Butucea and Y. I. Ingster. Detection of a sparse submatrix of a high-dimensional noisy matrix. *Bernoulli* **19** (5B) (2013) 2652–2688. MR3160567

[8] T. T. Cai, X. J. Jeng and J. Jin. Optimal detection of heterogeneous and heteroscedastic mixtures. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **73** (5) (2011) 629–662.

[9] T. T. Cai, J. Jin and M. G. Low. Estimation and confidence sets for sparse normal mixtures. *Ann. Statist.* **35** (6) (2007) 2421–2449. MR2382653

[10] T. T. Cai and Y. Wu. Optimal detection of sparse mixtures against a given null distribution. *IEEE Trans. Inform. Theory* **60** (4) (2014) 2217–2232. MR3181520

[11] H. Chernoff and E. Lander. Asymptotic distribution of the likelihood ratio test that a mixture of two binomials is a single binomial. *J. Statist. Plann. Inference* **43** (1–2) (1995) 19–40. MR1314126

[12] D. Donoho and J. Jin. Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.* **32** (3) (2004) 962–994. MR2065195

[13] H. Finner and M. Roters. Log-concavity and inequalities for chi-square, F and beta distributions with applications in multiple comparisons. *Statist. Sinica* **7** (1997) 771–787. MR1467456

[14] B. Garel. Recent asymptotic results in testing for mixtures. *Comput. Statist. Data Anal.* **51** (11) (2007) 5295–5304. MR2370872

[15] Y. I. Ingster and I. A. Suslina. *Nonparametric Goodness-of-Fit Testing Under Gaussian Models*. *Lecture Notes in Statistics* **169**, xiv+453. Springer-Verlag, New York, 2003. MR1991446

[16] Y. I. Ingster. Minimax detection of a signal for l^n-balls. *Math. Methods Statist.* **7** (4) (1998) 401–428. MR1680087

[17] B. Laurent, J. M. Loubes and C. Marteau. Non asymptotic minimax rates of testing in signal detection with heterogeneous variances. *Electron. J. Stat.* **6** (2012) 91–122. MR2879673

[18] B. Laurent, C. Marteau and C. Maugis-Rabusseau. Non-asymptotic detection of two-component mixtures with unknown mean. *Bernoulli* **22** (1) (2016) 242–274. MR3449782

[19] B. Laurent and P. Massart. Adaptive estimation of a quadratic functional by model selection. *Ann. Statist.* **28** (5) (2000) 1302–1338. MR1805785

[20] G. McLachlan and D. Peel. *Finite Mixture Models*. Wiley series in Probability and Statistics, New York, 2000. MR1789474

[21] N. Verzelen and E. Arias-Castro. Detection and feature selection in sparse mixture models. *Ann. Statist.* To appear. Available at arXiv:1405.1478.