Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2016, Vol. 52, No. 4, 1564–1591
DOI: 10.1214/15-AIHP689

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

Sharp detection of smooth signals in a high-dimensional sparse matrix with indirect observations

Cristina Butucea[a,c] and Ghislaine Gayraud[b,c]

[a]Université Paris-Est, LAMA (UMR 8050), UPEM, UPEC, CNRS F-77454, Marne-la-Vallée, France. E-mail: cristina.butucea@univ-mlv.fr
[b]Sorbonne universités, Université de Technologie de Compiègne, LMAC, CS 60 319, 60 203 Compiègne Cedex, France. E-mail: ghislaine.gayraud@utc.fr
[c]CREST-ENSAE, 3 Avenue Pierre Larousse, 92240 Malakoff, France

Received 5 May 2014; revised 21 March 2015; accepted 21 May 2015

Abstract. We consider a matrix-valued Gaussian sequence model, that is, we observe a sequence of high-dimensional $M \times N$ matrices of heterogeneous Gaussian random variables $x_{ij,k}$ for $i \in \{1, \ldots, M\}$, $j \in \{1, \ldots, N\}$, M and N tend to infinity and $k \in \mathbb{Z}$. For large $|k|$, the standard deviation of our observations is $\epsilon|k|^s$ for some $\epsilon > 0$, $\epsilon \to 0$ and a given $s \geq 0$, case that encompasses mildly ill-posed inverse problems.

We give separation rates for the detection of a sparse submatrix of size $m \times n$ (m and n tend to infinity, m/M and n/N tend 0) with active components. A component (i, j) is said active if the sequence $\{x_{ij,k}\}_k$ has mean $\{\theta_{ij,k}\}_k$ within a Sobolev ellipsoid of smoothness $\tau > 0$ and total energy $\sum_k \theta_{ij,k}^2$ larger than some r_ϵ^2. We construct a test procedure and compute rates that involve relationships between m, n, M, N and ϵ, such that the asymptotic total error probability tends to 0. We also show how these rates can be made adaptive to the size (m, n) of the submatrix under some constraints.

We prove corresponding lower bounds under additional assumptions on the relative size of the submatrix in the large matrix of observations. Our separation rates are sharp under further assumptions. Lower bounds for hypothesis testing problems mean that no test procedure can distinguish between the null hypothesis (no signal) and the alternative, i.e. the minimax total error probability for testing tends to 1.

Résumé. Nous considérons un modèle de suites de matrices de taille $M \times N$ dont les entrées sont des variables aléatoires Gaussiennes hétérogènes, $x_{ij,k}$, $i \in \{1, \ldots, M\}$, $j \in \{1, \ldots, N\}$, avec M et N qui tendent vers l'infini et $k \in \mathbb{Z}$. Pour $|k|$ grand, nous supposons l'écart-type de $x_{ij,k}$ de l'ordre de $\epsilon|k|^s$ pour $\epsilon > 0$ tel que $\epsilon \to 0$ et avec $s > 0$ connu; notre modèle permet donc d'inclure le cadre des problèmes inverses modérément mal-posés.

Nos résultats sont des vitesses de séparation dans le problème de détection d'une sous-matrice significative de taille $m \times n$, avec m et n qui tendent vers l'infini et parcimonieuse, c-à-d m/M et n/N tendent vers 0. Une composante (i, j) est dite active si la suite $\{x_{ij,k}\}_k$ a une espérance $\{\theta_{ij,k}\}_k$ qui appartient à une ellipsoide de Sobolev de régularité $\tau > 0$ et une énergie totale $\sum_k \theta_{ij,k}^2$ supérieure à r_ϵ^2. Nous construisons une procédure de test pour laquelle nous obtenons des vitesses de séparation impliquant des relations entre m, n, M, N et ϵ, de sorte que l'erreur totale de test tende vers 0. Nous montrons comment rendre ces vitesses de tests adaptatives en (m, n), la taille des sous-matrices significatives.

En faisant une hypothèse supplémentaire sur la taille relative des sous-matrices à détecter, nous prouvons les bornes inférieures correspondantes, ce qui assure qu'aucune procédure de test n'est capable de distinguer l'hypothèse nulle de l'alternative avec des vitesses « meilleures » que celles obtenues par notre procédure de test. Dans certains cas, nous obtenons des vitesses de séparation exactes.

MSC: 62H15; 60G15; 62G10; 62G20; 62C20

Keywords: Asymptotic minimax test; Detection boundary; Heterogeneous observations; Gaussian white noise model; High-dimensional data; Indirect observations; Inverse problems; Sharp rates; Sparsity

1. Introduction

We study here a collection of signals observed in Gaussian white noise models, doubly indexed by i from 1 to M and by j from 1 to N. We say that a component (i, j) is active if there is some signal at these coordinates, otherwise we assume that we observe only noise. We propose a procedure to test whether a smaller submatrix only contains active components, that is smooth signal with some given smoothness and significant energy (measured by its $\mathbb{L}_2$-norm). This step should be taken as a preliminary step to estimation methods designed for the sparse case.

Our model combines two problems with various applications to real-life data: on the one hand, cluster and bi-cluster detection in large matrices which are frequently used e.g. in genomics, signal theory, medical statistics and, on the other hand, sparse additive models which are very popular in machine learning.

Clusters' detection in spatial data was treated in the Bernoulli model by [25], and in the Gaussian model by [2] where the authors mention applications in biosurveillance, sensor arrays, digital images. Bi-clustering in large matrices are applied to genomics (DNA microarray data), biology (detecting groups of drugs and proteins that interact), computer science (detecting malware with similar signatures), marketing (detecting groups of clients with similar tastes for commercial products). These data can be functional (over time in biosurveillance or medical imaging, for example) and turned into signals, hence our model.

In other areas, signals are the object of interest and detection of activity in a sparse submatrix can arise naturally like in video surveillance, environment monitoring (transport of hazardous materials, detection of biological and chemical species). Sparse additive models are extensively used in learning theory, see [20] and references therein. These models correspond to the vector case. However, [20] mention that the more general functional ANOVA should allow interactions of order 2 and higher than 2, but that their computational complexity is too involved to be studied. Our problem can be used to detect sparse interactions of order 2 in the functional ANOVA model. The submatrix structure is only natural for the sparsity as the interactions propagate.

This problem can be stated equivalently in the Gaussian sequence model of coefficients of the signals (say Fourier coefficients), indexed by integer numbers k. We propose to deal with the Gaussian sequence model, as it is easier for our computations and discuss later on the alternative interpretation as signal detection. We also consider heterogeneous Gaussian observations in order to include the setup of indirect observations.

More precisely, we consider the following Gaussian sequence model

$$x_{ij,k} = \xi_{ij}\theta_{ij,k} + \epsilon\sigma_k\eta_{ij,k}, \quad i \in I = \{1, \dots, M\}, j \in J = \{1, \dots, N\}, k \in \mathbb{Z}, \tag{1.1}$$

where $\{\eta_{ij,k}\}_{i\in I, j\in J, k\in\mathbb{Z}}$ is a sequence of independent standard Gaussian random variables, $\sigma_k > 0$ and $\epsilon > 0$ are known. The $M \times N$-matrix $\xi = [\xi_{ij}]_{(i,j)\in I\times J}$, is deterministic (unknown) and has elements in $\{0, 1\}$.

In what follows, the standard deviations σ_k are supposed to be the same for all components of the matrix, that is they do not depend on (i, j) in $I \times J$. We assume throughout the paper that, for some fixed given $s \geq 0$,

$$\sigma_k \sim |k|^s, \quad \text{for large enough integer values of } |k|.$$

On the one hand, the case $s = 0$ reduces to the case of direct observations of the signal. In that case, we could generalize our results to unknown (but constant) variance σ. On the other hand, the case $s > 0$ corresponds to signals observed in inverse problems like convolution with some independent noise, tomography etc.

The polynomial behaviour of σ_k as k grows to infinity corresponds to mildly ill-posed inverse problems. We refer to Cavalier *et al.* [5] for more discussion on the relation between the sequence model with increasing variance and inverse problems in the Gaussian white noise model.

The matrix-valued sequence $\overline{\boldsymbol{\theta}} = [\xi_{ij}\{\theta_{ij,k}\}_{k\in\mathbb{Z}}]_{(i,j)\in I\times J}$ is the quantity of interest. We want to detect from observations in the model (1.1) whether there is only noise or whether there are active components in $\overline{\boldsymbol{\theta}}$, corresponding to (i, j) where $\xi_{ij} = 1$. When a component (i, j) is active, we assume that the corresponding sequence $\{\theta_{ij,k}\}_k$ belongs to a Sobolev ellipsoid and has significant total energy, i.e., $\{\theta_{ij,k}\}_k \in \Sigma(\tau, r_\epsilon)$, $\tau > 0, r_\epsilon > 0$, where

$$\Sigma(\tau, r_\epsilon) = \left\{\boldsymbol{\theta} \in l_2(\mathbb{Z}) : (2\pi)^{2\tau} \sum_{k\in\mathbb{Z}} |k|^{2\tau}\theta_k^2 \leq 1; \sum_{k\in\mathbb{Z}} \theta_k^2 \geq r_\epsilon^2\right\}. \tag{1.2}$$

In this paper, we assume that ξ has a specific structure, i.e., it belongs to

$$T_{M,N}(m,n) = \big\{\xi \text{ matrix of size } M \times N : \exists A_\xi \subseteq I, \#A_\xi = m \text{ and } \exists B_\xi \subseteq J, \#B_\xi = n \\ \text{such that } \xi_{ij} = \mathbb{1}\big((i,j) \in A_\xi \times B_\xi\big)\big\},$$

where the non-null elements form a submatrix with m rows and n columns (m and n are known) and with the notation $\#C$ for the cardinality of a set C. From now on, we shall denote by A_ξ and B_ξ those rows and columns where the matrix $\xi \in T_{M,N}(m,n)$ has non-null elements.

The testing problem of interest is the following

$$H_0 : \overline{\boldsymbol{\theta}} = 0,$$
$$H_1(\tau, r_\epsilon) : \overline{\boldsymbol{\theta}} \in \Theta_{M,N}(\tau, r_\epsilon, m, n),$$

where, for $\tau, r_\epsilon > 0$ and for m, n, M and N large, such that $m \le M$ and $n \le N$, we define

$$\Theta_{M,N}(\tau, r_\epsilon, m, n) = \big\{\overline{\boldsymbol{\theta}} = \big[\xi_{ij}\{\theta_{ij,k}\}_{k\in\mathbb{Z}}\big]_{(i,j)\in I\times J} : \xi \in T_{M,N}(m,n), \\ \text{and for all } (i,j) \in A_\xi \times B_\xi, \{\theta_{ij,k}\}_k \in \Sigma(\tau, r_\epsilon)\big\}.$$

The alternative hypothesis consists of matrices of size $M \times N$ containing mainly noise, except for elements in some submatrix of size $m \times n$ containing sequences of Fourier coefficients of signals with Sobolev smoothness τ and energy ($\mathbb{L}_2$ norm) significantly large (larger than r_ϵ).

The aim of this paper is to derive asymptotic detection boundaries, that is, asymptotic conditions allowing us to distinguish the hypotheses and separation rates, as defined later in Section 1.2, for alternatives $\Theta_{M,N}(\tau, r_\epsilon, m, n)$, and also to determine statistical procedures ψ which achieve these separation rates; such test procedures are said to be asymptotically minimax.

Remark 1.1. *We may also assume that the matrix ξ has entries either* 0 *or* 1, *such that $\sum_{(i,j)\in I\times J} \xi_{ij} = m \times n$. That means that we know the number of non-null elements of the matrix ξ since we know m and n but they can be found anywhere in the matrix. This case reduces exactly to the vector case previously studied by Gayraud and Ingster* [7] *under the sparsity condition that the number of active components mn satisfies $mn = (MN)^{1-b}$, where $b \in (0,1)$ corresponds to the sparsity index.*

Section 1.1 explains how this model is related to the multivariate Gaussian white noise model and how the inverse problem reduces to heterogeneous observations in our Gaussian sequence model. In Section 1.2 we give more notation and definitions.

1.1. *Sparse high-dimensional signal detection*

Let us see that the previous problem arises in some classical statistical models and hence, it has a different interpretation. When dealing with high-dimensional data, we model functions of many variables with additive models. For many situations where additive models are employed see Stone [23] and references therein. Let us consider the multivariate Gaussian white noise model

$$dX(t) = f(t)\,dt + \epsilon \cdot dW(t), \quad t \in [0,1]^d, d \in \mathbb{N}, \tag{1.3}$$

$\epsilon > 0$ and $W(t)$ is the Wiener process. When estimating f in a nonparametric model, the curse of dimensionality makes the rates exponentially slow for large dimension d. Additive models, where $f(t) = \sum_{j=1}^d f_j(t_j)$, $t_j \in [0,1]$ and $\int_0^1 f_j = 0$ for all j from 1 to d, are estimated with much faster rates, but the global estimation risk still grows in a linear way with d. It is assumed in Gayraud and Ingster [7] that the univariate signal functions f_j belong to a class $\mathcal{S}(\tau, r_\epsilon)$, i.e., they have Sobolev smoothness τ and total energy $\int_0^1 |f_j|^2$ larger than r_ϵ^2. A function f is Sobolev smooth if it belongs to $\mathbb{L}_2([0,1])$ such that $\int |\tilde{f}(u)|^2(2\pi|u|)^{2\tau}\,du \le 1$ (where $\tilde{f}$ is the characteristic function of a function f) and τ is called its smoothness parameter.

If we need to cope with very high dimension d, sparsity assumptions help to reduce the dimension. In Gayraud and Ingster [7], it was assumed that only d^{1-b} for some $0<b<1$ coordinates are significantly active, i.e., $f(t)=\sum_{j=1}^{d}\xi_j f_j(t_j)$, $\xi_j\in\{0,1\}$ for all j from 1 to d such that $\sum_j \xi_j = d^{1-b}$. They solved the following test problem:

$$H_0: \text{all } \xi_j = 0 \text{ (no signal is detected in data)},$$

$$H_1(\tau, r_\epsilon): \text{there exists } d^{1-b} \text{ values of } j \text{ where } \xi_j = 1 \text{ and } f_j \in \mathcal{S}(\tau, r_\epsilon).$$

Different sharp detection rates were obtained along the values of $0<b<1$, but the setup of heterogeneous variables was not studied.

In our paper, we assume a sparse matrix structure for our additive model:

$$f(t)=\sum_{i=1}^{M}\sum_{j=1}^{N}\xi_{ij} f_{ij}(t_{ij}), \quad t_{ij}\in[0,1] \text{ and } \xi\in T_{M,N}(m,n), \tag{1.4}$$

such that $\int_0^1 f_{ij}=0$ for all i, j. We call the component (i, j) active if $\xi_{ij}=1$ and, in that case, we suppose that the signal in that coordinate belongs to the class $\mathcal{S}(\tau, r_\epsilon)$.

Let us reduce the sparse additive model (1.3) such that (1.4) holds for our initial model. Consider $\{\varphi_k\}_{k\in\mathbb{Z}}$ the Fourier orthonormal basis of $\mathbb{L}_2[0,1]$ and recall that $\varphi_0\equiv 1$. Define the multivariate orthonormal family, for $t\in[0,1]^{M\times N}$,

$$\Phi_{ij,k}(t)=\varphi_k(t_{ij})\cdot\prod_{(l,h)\neq(i,j)}\varphi_0(t_{lh})=\varphi_k(t_{ij}).$$

Then, project the signal in (1.3) on these functions:

$$\begin{aligned} x_{ij,k} &:= \int_{[0,1]^{M\times N}} \Phi_{ij,k}(t)\,dX(t) \\ &= \int_{[0,1]^{M\times N}} \Phi_{ij,k}(t) f(t)\,dt + \epsilon\cdot\int_{[0,1]^{M\times N}} \Phi_{ij,k}(t)\,dW(t) \\ &= \xi_{ij}\int_0^1 \varphi_k(t_{ij}) f_{ij}(t_{ij})\,dt_{ij} + \epsilon\cdot\eta_{ij,k}, \end{aligned}$$

where $\{\eta_{ij,k}\}$ are i.i.d. standard Gaussian random variables.

We get our initial model for $\theta_{ij,k}=\int_0^1\varphi_k f_{ij}$ and $\sigma_k\equiv 1$. Indeed, following Tsybakov [24], we know that f_{ij} belongs to $\mathcal{S}(\tau, r_\epsilon)$ if and only if $\{\theta_{ij,k}\}_k$ belongs to $\Sigma(\tau, r_\epsilon)$. Then, our test problem can be written:

$$H_0: \text{ all } \xi_{ij}=0 \text{ (no signal is detected in data)},$$

$$H_1(\tau, r_\epsilon): \text{ there exists } \xi\in T_{M,N}(m,n) \text{ and for } \xi_{ij}=1 \text{ it holds that } f_{ij}\in\mathcal{S}(\tau, r_\epsilon),$$

i.e., there exists a matrix ξ in $T_{M,N}(m,n)$ such that the signal in active coordinates (i, j) has Sobolev smoothness τ and total energy larger than r_ϵ^2.

The variance $\epsilon^2\sigma_k^2$ of our observations is allowed to increase with k, since $\sigma_k\sim|k|^s$, $s\ge 0$. Indeed, let us suppose that our additive model is observed in an inverse problem. That means that we observe

$$dX(t)=Kf(t)\,dt+\epsilon\cdot dW(t), \quad t=[t_{ij}]_{i,j}\in[0,1]^{M\times N} \tag{1.5}$$

for some linear operator K, with f given as in (1.4) and such that $\int_0^1 Kf_{ij}=0$.

Take, for example, the convolution model. In this case, the signal is observed with an additive independent noise having density g. Then the operator K acts as a convolution operator with the density g and writes $Kf(t) = \int f(t-u)g(u)\,du$.

We suppose that K^*K is a compact operator having eigenvalues σ_k^{-2} decreasing polynomially to 0 as k tends to infinity. This corresponds to mildly ill-posed inverse problems. Whereas, in the case of well-posed inverse problems, $\sigma_k^2 \le \sigma^2$ form a bounded sequence.

Then, we consider a singular value decomposition of K, that is families of orthonormal functions $\{\varphi_k\}_k$ and $\{\psi_k\}_k$ such that $K\varphi_k = \sigma_k^{-1}\psi_k$ and $K^*\psi_k = \sigma_k^{-1}\varphi_k$. Therefore, let $\psi_k \equiv 1$ and $\Psi_{ij,k}(t) = \psi_k(t_{ij})$, and project (1.5) on this family:

$$\begin{aligned} y_{ij,k} &:= \sum_{l=1}^{M}\sum_{h=1}^{N} \xi_{lh} \int_{[0,1]^{M\times N}} \Psi_{ij,k}(t) K f_{lh}(t_{lh})\,dt_{lh} + \epsilon \cdot \int_{[0,1]^{M\times N}} \Psi_{ij,k}(t)\,dW(t) \\ &= \xi_{ij} \int_0^1 \psi_k(u) K f_{ij}(u)\,du + \epsilon \cdot \eta_{ij,k}. \end{aligned}$$

Note, moreover, that $\int_0^1 \psi_k \cdot K f_{ij} = \int_0^1 K^*\psi_k \cdot f_{ij} = \sigma_k^{-1}\int_0^1 \varphi_k \cdot f_{ij} = \sigma_k^{-1}\theta_{ij,k}$. Then, let $x_{ij,k} = \sigma_k y_{ij,k}$ to get the model (1.1).

Note that Butucea and Ingster [4] studied the particular case where $\theta_{ij,k} = a\mathbb{1}(k=0)$ and the variance of the noise is a given fixed σ. The asymptotic rates for testing were given in terms of n, m, N and M. Here, we replace the constant elements with arbitrary signals having a given amount of smoothness. Moreover, we add here the case of heterogeneous variables which include mildly ill-posed inverse problems.

1.2. *Notation and definitions*

Denote by $\mathbb{P}_0$ and $\mathbb{P}_{\overline{\boldsymbol{\theta}}}$ the distributions under the null and the alternative, respectively. Denote also by $\mathbb{E}_0$, Var_0 and $\mathbb{E}_{\overline{\boldsymbol{\theta}}}$, $\mathrm{Var}_{\overline{\boldsymbol{\theta}}}$ the expected values and variances with respect to $\mathbb{P}_0$ and $\mathbb{P}_{\overline{\boldsymbol{\theta}}}$, respectively. Set $\boldsymbol{\theta}_{ij} = \{\theta_{ij,k}\}_{k\in\mathbb{Z}}$; indices of probabilities, expectations or variances which are expressed in terms of non-overlined subsequences of $\boldsymbol{\theta}$ mean that they correspond to active components.

For any test procedure ψ, that is, any measurable function with respect to the observations, taking values in [0, 1], set $\omega(\psi) = \mathbb{E}_0(\psi)$ its type I error probability and $\beta(\psi, \Theta_{M,N}(\tau, r_\epsilon, m, n)) = \sup_{\overline{\boldsymbol{\theta}}\in\Theta_{M,N}(\tau,r_\epsilon,m,n)} \mathbb{E}_{\overline{\boldsymbol{\theta}}}(1-\psi)$ its maximal type II error probability over the set $\Theta_{M,N}(\tau, r_\epsilon, m, n)$. Let us denote by

$$\gamma\bigl(\psi, \Theta_{M,N}(\tau, r_\epsilon, m, n)\bigr) = \omega(\psi) + \beta\bigl(\psi, \Theta_{M,N}(\tau, r_\epsilon, m, n)\bigr)$$

the total error probability of ψ and denote by γ the minimax total error probability over $\Theta_{M,N}(\tau, r_\epsilon, m, n)$ which is defined by

$$\gamma := \gamma\bigl(\Theta_{M,N}(\tau, r_\epsilon, m, n)\bigr) = \inf_{\psi} \gamma\bigl(\psi, \Theta_{M,N}(\tau, r_\epsilon, m, n)\bigr),$$

where the infimum is taken over all test procedures. We can not distinguish H_0 and $H_1(\tau, r_\epsilon)$ if $\gamma \to 1$ and we say that we can distinguish the hypotheses if there exists ψ such that $\gamma(\psi, \Theta_{M,N}(\tau, r_\epsilon, m, n)) \to 0$.

By (asymptotic) separation rates or minimax rates of testing, we mean a sequence $\tilde{r_\epsilon}$ such that

$$\begin{cases} \gamma \to 1 & \text{if } \frac{r_\epsilon}{\tilde{r_\epsilon}} \to 0, \\ \gamma(\psi, \Theta_{M,N}(\tau, r_\epsilon, m, n)) \to 0 & \text{if } \frac{r_\epsilon}{\tilde{r_\epsilon}} \to +\infty. \end{cases}$$

By (asymptotic) sharp separation rates or sharp minimax rates of testing, we mean a sequence $\tilde{r_\epsilon}$ such that

$$\begin{cases} \gamma \to 1 & \text{if } \limsup \frac{r_\epsilon}{\tilde{r_\epsilon}} < 1, \\ \gamma(\psi, \Theta_{M,N}(\tau, r_\epsilon, m, n)) \to 0 & \text{if } \liminf \frac{r_\epsilon}{\tilde{r_\epsilon}} > 1. \end{cases}$$

The asymptotics for model (1.1) are given when $\epsilon \to 0$ and, as we are mainly interested in high-dimensional settings, when

$$m, n, M \text{ and } N \to +\infty, \qquad p = \frac{m}{M} \to 0, \qquad q = \frac{n}{N} \to 0. \tag{1.6}$$

Here and later asymptotics and symbols o, O, $\sim$ and $\asymp$ are considered under $\epsilon \to 0$ and m, n, M and N such that (1.6) holds. Recall that, given sequences of real numbers u and real positive numbers v, we say that they are asymptotically equivalent, $u \sim v$, if $\lim u/v = 1$. Moreover, we say that the sequences are asymptotically of the same order, $u \asymp v$, if there exist two constants $0 < c \le C < \infty$ such that $c \le \liminf u/v$ and $\limsup u/v \le C$.

Nonparametric tests in the minimax setup were introduced by Ingster [8], [9] and [10] and intensively studied ever since. In the context of inverse problems, minimax testing was considered by Butucea [3] in the density model, Ingster *et al.* [13] in the Gaussian sequence model and [12] for a problem related to the Radon transform. A non-asymptotic study of minimax rates in inverse problems was given in Laurent *et al.* [19]. Other minimax testing rates in multivariate setups were given by Ingster and Stepanova [14,15] and Laurent *et al.* [18].

Detection of sparse vectors was treated by Ingster [11], Ingster and Suslina [16] and Donoho and Jin [6]. Detection of sparse matrices was studied by Butucea and Ingster [4]. As we already mentioned, Gayraud and Ingster [7] consider multivariate functions depending only on a small number of coordinates. In this paper, we assume that the sparse coordinates have a submatrix structure. Moreover, in this setup we include the setup of inverse problems. Detection of sparse objects has not been considered before in inverse problems to the best of our knowledge.

The plan of the paper is as follows. In Section 2 we define the test procedure and give sufficient conditions such that its total error probability tends to 0. The construction of our test procedure involves solving an optimization problem, which determines the construction of the minimax test procedure. Section 3 presents the lower bounds for our problem. In Section 4 we give related results: first, particular cases where our separation rates write in a simpler form, then we make our procedure universal with respect to the separation parameter r_ϵ and, finally, in Section 4.3 we show that the separation rates are still attained uniformly over a set of values (m, n). Proofs are given in Section 5 and the Appendix.

2. Testing procedures and their asymptotic behaviour

Consider the following family of weighted χ^2-type statistics: for (i, j) in $I \times J$

$$t_{ij,w} = \sum_{k \in \mathbb{Z}} w_k \left(\left(\frac{x_{ij,k}}{\epsilon \sigma_k} \right)^2 - 1 \right), \tag{2.1}$$

where $(w_k)_k$ is a sequence of weights such that $w_k \ge 0$ for all $k \in \mathbb{Z}$ and $\sum_{k \in \mathbb{Z}} w_k^2 = 1/2$.

In order to define the weights $\{w_k^\star\}_{k \in \mathbb{Z}}$ that will appear in the optimal test procedure, we solve the following extremal problem. Recall that $\Sigma(\tau, r_\epsilon)$ denotes the Sobolev ellipsoid defined in (1.2), with $\tau > 0$ and $r_\epsilon > 0$, and $\{\sigma_k\}_{k \in \mathbb{Z}}$ is a sequence of positive real numbers. We define the sequences $\{w_k^\star\}_{k \in \mathbb{Z}}$ and $\{\theta_k^\star\}_{k \in \mathbb{Z}}$ as solutions to the following optimization program:

$$\sum_{k \in \mathbb{Z}} w_k^\star \left(\frac{\theta_k^\star}{\epsilon \sigma_k} \right)^2 = \sup_{\{(w_k)_k \in l^2(\mathbb{Z}) : w_k \ge 0; \sum_k w_k^2 = \frac{1}{2}\}} \inf_{\{\theta_k\}_k \in \Sigma(\tau, r_\epsilon)} \sum_{k \in \mathbb{Z}} w_k \left(\frac{\theta_k}{\epsilon \sigma_k} \right)^2. \tag{2.2}$$

Let us denote by $a(r_\epsilon) := \sum_{k \in \mathbb{Z}} w_k^\star (\theta_k^\star / (\epsilon \sigma_k))^2$, the value of the optimization problem (2.2) at the optimal point.

Let us discuss heuristically why we need to solve this problem, before giving the solution. Note that under the null hypothesis our statistic becomes $t_{ij,w} = \sum_{k \in \mathbb{Z}} w_k (\eta_{ij,k}^2 - 1)$ and it is a standard random variable (due to the normalization $\sum_{k \in \mathbb{Z}} w_k^2 = 1/2$). Under the alternative,

$$\mathbb{E}_{\boldsymbol{\theta}_{ij}} (t_{ij,w}) = \sum_{k \in \mathbb{Z}} w_k \left(\frac{\theta_{ij,k}}{\epsilon \sigma_k} \right)^2. \tag{2.3}$$

In order to distinguish the alternative from the null at best, we need to consider the worst parameter $\boldsymbol{\theta}_{ij}$ under the alternative and then maximize over possible weights $w_k \geq 0$ verifying the normalization constraints $\sum_k w_k^2 = 1/2$.

Proposition 2.1. *Let $\{\sigma_k\}_{k\in\mathbb{Z}}$ be a sequence of positive real numbers such that $\sigma_k \sim |k|^s$ as $|k|$ large enough, for a given $s > 0$. Then, the optimization problem* (2.2) *has the following solution*:

$$(\theta_k^\star)^2 = v\sigma_k^4\sqrt{2}\left(1-\left(\frac{|k|}{T}\right)^{2\tau}\right)_+ \quad \text{and} \quad w_k^\star = \frac{(\theta_k^\star)^2}{2\epsilon^2\sigma_k^2 a(r_\epsilon)},$$

where $(x)_+ = \max(0, x)$, with

$$T \sim \left(\frac{\kappa_1}{\kappa_2}\right)^{1/(2\tau)} r_\epsilon^{-1/\tau}, \qquad v = \frac{1}{\kappa_1}\left(\frac{\kappa_2}{\kappa_1}\right)^{(4s+1)/(2\tau)} r_\epsilon^{2+(4s+1)/\tau} \quad \text{and} \quad a(r_\epsilon) \sim c(\tau,s)\epsilon^{-2} r_\epsilon^{2+(4s+1)/(2\tau)},$$

where the asymptotics are taken as $k \to \infty$ and as $r_\epsilon \to 0$ and the constants are given by

$$c(\tau,s)^2 = 2\left(\frac{\kappa_1}{\kappa_2}\right)^{-(4s+1)/(2\tau)} \frac{\kappa_3}{\kappa_1^2}, \qquad \kappa_1 = \frac{4\sqrt{2}\tau}{(4s+1)(4s+2\tau+1)},$$

$$\kappa_2 = \frac{4\sqrt{2}\tau(2\pi)^{2\tau}}{(4s+2\tau+1)(4s+4\tau+1)} \quad \text{and} \quad \kappa_3 = \frac{1}{4s+1} - \frac{2}{4s+2\tau+1} + \frac{1}{4s+4\tau+1}.$$

Moreover, we have $\sup_k w_k^\star \leq r_\epsilon^{1/(2\tau)} \to 0$.

The proof of Proposition 2.1 is postponed to the Appendix. Note that $w^\star = \{w_k^\star\}_{k\in\mathbb{Z}}$ and $\theta^\star = \{\theta_k^\star\}_{k\in\mathbb{Z}}$ check the constraints in (2.2), that is, $\sum_k (w_k^\star)^2 = \frac{1}{2}$, $\sum_k (\theta_k^\star)^2 = r_\epsilon^2(1+o(1))$ and $\sum_k (2\pi k)^{2\tau}(\theta_k^\star)^2 = 1 + o(1)$, as $r_\epsilon \to 0$. It is worthwhile to note that, due to Proposition 2.1 and relation (2.3), we have

$$\frac{1}{2}\sum_k \frac{(\theta_k^\star)^4}{\epsilon^4\sigma_k^4} = a^2(r_\epsilon), \tag{2.4}$$

$$\inf_{\boldsymbol{\theta}_{ij}\in\Sigma(\tau,r_\epsilon)} \mathbb{E}_{\boldsymbol{\theta}_{ij}}(t_{ij,w^\star}) = a(r_\epsilon) \tag{2.5}$$

and note also that the sequences $w^\star$ and $\theta^\star$ have a finite number T of non-null elements, but T grows to infinity as $r_\epsilon \to 0$.

Define the test procedures,

$$\psi^{\chi^2} = \mathbb{1}\left(t^{\chi^2} > H\right), \quad \text{with } t^{\chi^2} = \frac{1}{\sqrt{MN}} \sum_{(i,j)\in I\times J} t_{ij,w^\star}, \tag{2.6}$$

$$\psi^{\text{scan}} = \mathbb{1}\left(t^{\text{scan}} > K\right), \quad \text{with } t^{\text{scan}} = \max_{\xi\in T_{M,N}(m,n)} \frac{1}{\sqrt{mn}} \sum_{(i,j)\in A_\xi\times B_\xi} t_{ij,w^\star}, \tag{2.7}$$

where $K^2 = 2(1+\delta)\log(\binom{N}{n}\binom{M}{m})$ for some small $\delta > 0$ and H is a positive number, depending on s and τ, on ϵ, m, n, M and N, to be correctly chosen in the following theorem.

Under the assumption (1.6), we can check that

$$\log\left(\binom{M}{m}\cdot\binom{N}{n}\right) \sim m\cdot\log\left(p^{-1}\right) + n\cdot\log\left(q^{-1}\right).$$

The following theorem gives the upper bounds for the testing rates of the previously defined procedures. We denote by Φ the c.d.f. of a standard Gaussian distribution.

Theorem 2.1. *Assume* (1.6). *Suppose that* $r_\epsilon \to 0$ *and recall that*

$$a(r_\epsilon) \sim c(\tau, s)\epsilon^{-2} r_\epsilon^{2+(4s+1)/(2\tau)}.$$

1. *The linear test statistics* ψ^{χ^2} *defined by* (2.6) *has the following properties.*
   Type I error probability: *if* $H \to \infty$, *then* $\omega(\psi^{\chi^2}) = \Phi(-H) + o(1)$.
   Maximal type II error probability: *if*
   $$a^2(r_\epsilon)mnpq \to +\infty, \tag{2.8}$$
   choose H *such that* $H \le c \cdot a(r_\epsilon)\sqrt{mnpq}$, *for some* $0 < c < 1$, *then* $\beta(\psi^{\chi^2}, \Theta_{M,N}(\tau, r_\epsilon, m, n)) = o(1)$.
2. *The scan test statistic* ψ^{scan} *defined by* (2.7) *has the following properties.*
   Suppose that $K^2 r_\epsilon^{1/\tau}/(mn) = o(1)$.
   Type I error probability: $\omega(\psi^{\text{scan}}) = o(1)$.
   Maximal type II error probability: *if*
   $$\liminf \frac{a^2(r_\epsilon)mn}{2(m \cdot \log(p^{-1}) + n \cdot \log(q^{-1}))} > 1, \tag{2.9}$$
   then $\beta(\psi^{\text{scan}}, \Theta_{M,N}(\tau, r_\epsilon, m, n)) = o(1)$.

Let us note that the condition $K^2 r_\epsilon^{1/\tau}/(mn) = o(1)$ is pretty mild on the size of the submatrix. Indeed, as we assume that $r_\epsilon \to 0$, it comes down to assume that $\log(p^{-1})/n + \log(q^{-1})/m = o(r_\epsilon^\tau)$. This condition allows us to tune K at its smallest possible value by Lemma 5.1 so that the type II error probability tends to 0 under the sharp condition (2.9).

As a consequence of Theorem 2.1, we have the following result.

Corollary 2.1. *Assume* (1.6). *Suppose that* $r_\epsilon \to 0$. *Consider* ψ *the test procedure which combines* ψ^{χ^2} *and* ψ^{scan} *as follows*

$$\psi = \max(\psi^{\chi^2}, \psi^{\text{scan}}).$$

The test procedure ψ *with* H *and* K *chosen as in Theorem* 2.1 *is such that* $\gamma(\psi, \Theta_{M,N}(\tau, r_\epsilon, m, n)) = o(1)$ *as soon as either* (2.8) *or* (2.9) *hold.*

3. Minimax total error probability

We prove here optimality results for the rates attained by the previous test procedure ψ. However, the optimality is attained under additional hypothesis requiring an 'almost' squared matrix in the sense that the relative sizes of the submatrix should be of the same order in both directions (rows and columns sizes). More precisely, these additional hypotheses are that

$$\frac{\log\log(p^{-1})}{\log(q^{-1})} \to 0, \qquad \frac{\log\log(q^{-1})}{\log(p^{-1})} \to 0, \tag{3.1}$$

$$m \cdot \log(p^{-1}) \asymp n \cdot \log(q^{-1}) \tag{3.2}$$

and

$$\frac{m \cdot \log(p^{-1}) + n \cdot \log(q^{-1})}{mn} = o(1)\epsilon^{-2/(2\tau+2s+1)}. \tag{3.3}$$

Theorem 3.1. *Assume* (1.6) *and* (3.1)–(3.3). *If* r_ϵ *is such that the following conditions are satisfied*

$$a^2(r_\epsilon)\cdot mnpq \to 0, \tag{3.4}$$

$$\limsup \frac{a^2(r_\epsilon)\cdot mn}{2(m\cdot \log(p^{-1}) + n\cdot \log(q^{-1}))} < 1, \tag{3.5}$$

then $\inf_\psi \gamma(\psi, \Theta_{M,N}(\tau, r_\epsilon, m, n)) \to 1$.

The proof of Theorem 3.1 is given in Section 5.3.

Remark 3.1. *Theorems* 2.1 *and* 3.1 *together say that, under assumptions* (1.6), (3.1), (3.2) *and* (3.3) *our test procedure* ψ *in Corollary* 2.1 *is asymptotically minimax, it achieves the lower bounds.*

Let us insist on the complementarity of the conditions on r_ϵ *such that, on the one hand, the test procedure* ψ *has total error probability tending to* 0 *and that, on the other hand, no test procedure can distinguish the two hypotheses. Our results provide separation rates between these cases and they are defined through* $a(r_\epsilon)$. *Indeed, the detection boundary* $a(\tilde{r_\epsilon})$ *satisfies the following relations*

$$a^2(\tilde{r_\epsilon})\cdot mnpq \asymp 1, \qquad a^2(\tilde{r_\epsilon})\cdot mn \sim 2\bigl(m\cdot \log\bigl(p^{-1}\bigr) + n\cdot \log\bigl(q^{-1}\bigr)\bigr).$$

Therefore, the detection boundary can be written

$$a(\tilde{r_\epsilon}) \asymp \min\left\{\frac{1}{\sqrt{mnpq}}, \sqrt{\frac{2(m\cdot \log(p^{-1}) + n\cdot \log(q^{-1}))}{mn}}\right\}, \tag{3.6}$$

and they are sharp if $\frac{2(m\cdot\log(p^{-1})+n\cdot\log(q^{-1}))}{mn} \le \frac{1}{mnpq}$, *that is when the scan statistic detects. By Proposition* 2.1, *we have that* $a(\tilde{r_\epsilon}) \sim c(\tau, s)\epsilon^{-2}(\tilde{r_\epsilon})^{2+(4s+1)/(2\tau)}$ *as* $\epsilon \to 0$ *and* $\tilde{r_\epsilon} \to 0$ *and it implies that* $\epsilon^2 a(\tilde{r_\epsilon}) \to 0$.

Remark 3.2. *Note that for the test procedure* ψ^{χ^2} *we show separation rates, while for* ψ^{scan} *we show sharp separation rates. This is intrinsic to the model. More refined results can be stated in order to get sharp separation rates for the linear procedure. In particular, following* [13], *which is based on* [17], *we have* (*for one-dimensional signal*) *that, if*

$$a^2(r_\epsilon) \asymp 1 \quad \textit{and} \quad \sup_k w_k = o(1),$$

then the minimax total error probability is such that

$$\gamma\bigl(\Theta_{1,1}(\tau, r_\epsilon, 1, 1)\bigr) = 2\Phi\left(-\frac{1}{2}a(r_\epsilon)\right) + o(1).$$

Obtaining analogous results for the sparse matrix case is beyond the scope of our paper.

4. Related results

In this section, we discuss various results. We start by giving the rates in some particular cases, some of them are already known. Then we go back to our testing procedure and show how to make it universal with respect to the rate r_ϵ. In the penultimate part, we make the procedure adaptive to the size of the submatrix (m, n) varying in some set $\mathcal{K}_{M,N}$, without loss in the separation rate, under some constraints on the set. Finally, we add a comment on how we could implement our test procedures.

4.1. *Particular cases*

In this section we discuss how our results connect with known cases and how they write in particular examples.

One-dimensional signal. We can recover from these results the separation rates for one-dimensional sequences (i.e. $M = N = m = n = 1$). In this case (3.6) requires that $a(\tilde{r_\epsilon})$ is asymptotically constant, which means $\tilde{r_\epsilon} \sim (c^{-1}(\tau, s)\epsilon^2)^{2\tau/(4\tau+4s+1)}$ and that is the minimax rate for testing one-dimensional signal with Sobolev smoothness τ, see Ingster *et al.* [13].

Polynomially sparse submatrix. Let us consider the particular case where $m = M^{1-b_1}$ and $n = N^{1-b_2}$ with the sparsity indices b_1 and b_2 in $(0, 1)$. Suppose that $M^C \asymp N$ with some positive C, then assumption (3.2) is satisfied provided that $(1 - b_1) = C(1 - b_2)$. The particular choice of $C = 1$ is discussed in details later.

For any positive C, note that assumption (1.6) is reduced to $M \to \infty$. Moreover, assumption (3.1) is trivially satisfied and condition (3.3) becomes $M^{b_1-1} = o(\epsilon^{-2/(2\tau+2s+1)})$. Then, the detection boundary, namely $a(\tilde{r}_\epsilon)$ in (3.6), satisfies

$$a^2(\tilde{r}_\epsilon) \asymp \min\{M^{4b_1-2}, 2(b_1 + Cb_2)\log(M)M^{b_1-1}\}.$$

Due to Proposition 2.1, we can deduce separation rates which are reported in Table 1; those separation rates are different according to the values of (b_1, b_2) in the open square $]0, 1[\times]0, 1[$. It is worthwhile to note that for $0 < b_1 < 1/2$ and $0 < b_2 < \min(1, (1 - 2b_1)/C)$, the separation rate corresponds to the one obtained in the M^2-dimensional Gaussian white noise model when the sparsity is moderate (see Gayraud and Ingster [7]). This seems reasonable as the linear statistic detects the submatrix in those cases and its behaviour does not depend on the vector or matrix setup. Note that our results generalize the previous rates for the more general model including ill-posed inverse problems.

Squared polynomially sparse submatrix. Now, the choice $C = 1$ in the previous case, which leads to $b_1 = b_2 := b$ and hence $M \asymp N$, corresponds to the case of a sparse squared submatrix in a squared matrix. In this case, the separation rates, reported in Table 2 are different according to the values of b. Here the cut-off is $b = 1/3$ and again the separation rate corresponds to the one obtained in the M^2-dimensional Gaussian white noise model when the sparsity is moderate (see Gayraud and Ingster [7]).

Sparse high-dimensional vector. Let us assume $n = N = 1$ and $m = M^{1-b}$, with $b \in (0, 1)$ the sparsity index. It is obvious that conditions (3.1) and (3.2) are not satisfied, but the upper bounds hold and they give the separation rates

$$\tilde{r}_\epsilon \asymp \min\{\left(c^{-2}(\tau, s)\epsilon^4 M^{2b-1}\right)^{\tau/(4\tau+4s+1)}, \left(2bc^{-2}(\tau, s)\epsilon^4 \log(M)\right)^{\tau/(4\tau+4s+1)}\}.$$

Note that the rate is $(c^{-2}(\tau, s)\epsilon^4 M^{2b-1})^{\tau/(4\tau+4s+1)}$ when $b < 1/2$ and it is $(2bc^{-2}(\tau, s)\epsilon^4 \log(M))^{\tau/(4\tau+4s+1)}$ when $b > 1/2$. Again, in the moderately sparse case we find the separation rates of Gayraud and Ingster [7] in the more general setup of inverse problems. However, the highly-sparse vector case should be reconsidered in the case of inverse problems.

Table 1
Separation rates $\tilde{r}_\epsilon$

	$b_1 \in (0, 1/2)$	$b_1 \in [1/2, 1)$
$0 < b_2 < \min(1, \frac{1-2b_1}{C})$	$(c^{-2}(\tau, s)\epsilon^4 M^{4b_1-2})^{\tau/(4\tau+4s+1)}$	$(\frac{2(b_1+Cb_2)\epsilon^4}{c^2(\tau,s)} M^{b_1-1}\log(M))^{\tau/(4\tau+4s+1)}$
$\frac{1-2b_1}{C} < b_2 < 1$	$(\frac{2(b_1+Cb_2)\epsilon^4}{c^2(\tau,s)} M^{b_1-1}\log(M))^{\tau/(4\tau+4s+1)}$	$(\frac{2(b_1+Cb_2)\epsilon^4}{c^2(\tau,s)} M^{b_1-1}\log(M))^{\tau/(4\tau+4s+1)}$

Table 2
Detection boundary

	$b \in (0, 1/3]$	$b \in (1/3, 1)$
$a^2(\tilde{r}_\epsilon)$	M^{4b-2}	$4b\log(M)M^{b-1}$
$\tilde{r}_\epsilon$	$(c^{-2}(\tau, s)\epsilon^4 M^{4b-2})^{\tau/(4\tau+4s+1)}$	$(4bc^{-2}(\tau, s)\epsilon^4 M^{b-1}\log(M))^{\tau/(4\tau+4s+1)}$

4.2. *Universal test procedure*

Recall that the test statistics t^{χ^2} and t^{scan} depend on $t_{ij,w^\star}$ and hence on r_ϵ through the sequence $w^\star$ (see Proposition 2.1); it means that t^{χ^2} as well as t^{scan} are different for two distinct r_ϵ (given by the alternative hypotheses). One may be interested in dealing with test statistics free of r_ϵ in a range where either (2.8) or (2.9) hold. In this part, we describe such a procedure. Denote by $\tilde{r}_\epsilon^{\chi^2}$ and $\tilde{r}_\epsilon^{\text{scan}}$

$$\tilde{r}_\epsilon^{\chi^2} = (mnpq)^{-1/2}\epsilon^2 c^{-1}(\tau,s) \quad \text{and} \quad \tilde{r}_\epsilon^{\text{scan}} = \sqrt{2}\left(\frac{m\log p^{-1} + n\log q^{-1}}{mn}\right)^{-1/2}\epsilon^2 c^{-1}(\tau,s),$$

where $c(\tau,s)$ is defined in Proposition 2.1. Then define the test procedures $\tilde{\psi}^{\chi^2}$ and $\tilde{\psi}^{\text{scan}}$ as follows:

$$\tilde{\psi}^{\chi^2} = \mathbb{1}\big(\tilde{t}^{\chi^2} > \tilde{H}\big), \quad \text{with } \tilde{t}^{\chi^2} = \frac{1}{\sqrt{MN}}\sum_{(i,j)\in I\times J} t_{ij,w^*(\tilde{r}_\epsilon^{\chi^2})},$$

$$\tilde{\psi}^{\text{scan}} = \mathbb{1}\big(\tilde{t}^{\text{scan}} > K\big), \quad \text{with } \tilde{t}^{\text{scan}} = \max_{\xi\in T_{M,N}(m,n)} \frac{1}{\sqrt{mn}} \sum_{(i,j)\in A_\xi\times B_\xi} t_{ij,w^*(\tilde{r}_\epsilon^{\text{scan}})},$$

where $\tilde{H}$ is a positive constant such that $\tilde{H}\to\infty$ and recall that $K^2 = 2(1+\delta)\log\big(\binom{N}{n}\binom{M}{m}\big)$ for small positive δ.

Lemma 3.1 in Ingster and Suslina [17] implies that $\tilde{t}^{\chi^2}$ is asymptotically standard Gaussian under H_0 and hence the type I error probability of $\tilde{\psi}^{\chi^2}$ is asymptotically zero. Acting exactly as for ψ^{scan}, the type I error probability of $\tilde{\psi}^{\text{scan}}$ can be proved to vanish asymptotically under the condition $K(\tilde{r}_\epsilon^{\text{scan}})^{1/(2\tau)}/\sqrt{mn} = o(1)$.

To control the maximal type II error probability of $\tilde{\psi}^{\chi^2}$ and $\tilde{\psi}^{\text{scan}}$ over the set $\Theta_{M,N}(\tau, r_\epsilon, m, n)$, when r_ϵ satisfied either (2.8) or (2.9) reduces to bound from below, uniformly over $\Sigma(\tau, r_\epsilon)$, the term $\mathbb{E}_{\boldsymbol{\theta}_{ij}}(t_{ij,w^\star(\tilde{r}_\epsilon)})$, where $\tilde{r}_\epsilon$ stands for either $\tilde{r}_\epsilon^{\chi^2}$ or $\tilde{r}_\epsilon^{\text{scan}}$.

Note that due to Proposition 2.1, relations (2.8) and (2.9) are respectively equivalent to $r_\epsilon/\tilde{r}_\epsilon^{\chi^2}\to+\infty$ and $\liminf r_\epsilon/\tilde{r}_\epsilon^{\text{scan}} > 1$. Without loss of generality, let us suppose that r_ϵ is equal to $B\tilde{r}_\epsilon$, with $B>1$. Therefore, it suffices to bound from below $\mathbb{E}_{\boldsymbol{\theta}_{ij}}(t_{ij,w^\star(\tilde{r}_\epsilon)})$ uniformly over $\Sigma(\tau, B\tilde{r}_\epsilon)$ under the condition $B\to\infty$ and $B>1$ respectively. Now fix $\tilde{H} = B/2$.

Applying Proposition 4.1 in Gayraud and Ingster [7] gives that for any $\boldsymbol{\theta}_{ij}\in\Sigma(\tau, B\tilde{r}_\epsilon)$ with $B\ge 1$, one has

$$\mathbb{E}_{\boldsymbol{\theta}_{ij}}(t_{ij,w^\star(\tilde{r}_\epsilon)}) \ge B^2 a(\tilde{r}_\epsilon),$$

which leads to prove that the type II error vanishes asymptotically since $B-\tilde{H}\to\infty$ and $2(m\log(p^{-1}) + n\log(q^{-1}))(B^2-(1+\delta))\to\infty$ for δ small enough.

4.3. *Adaptation*

We shall consider here adaptation of our test procedure to the size of the submatrix (n,m).

Following Butucea and Ingster [4], we introduce a set of indices $\mathcal{K}_{M,N}\subseteq I\times J$ such that the following conditions hold:

$$\sup_{(m,n)\in\mathcal{K}_{M,N}}\left(\frac{1}{m}+\frac{1}{n}+\frac{m}{M}+\frac{n}{N}\right)\to 0 \quad \text{and} \quad \sup_{(m,n)\in\mathcal{K}_{M,N}}\left(\frac{\log M}{m\log(p^{-1})}+\frac{N}{n\log(q^{-1})}\right)\to 0.$$

For each $(m,n)\in\mathcal{K}_{M,N}$, consider the same statistic ψ^{χ^2} as defined in (2.6), but a slightly modified version of (2.7), namely the test

$$\tilde{\psi}^{\text{scan}}_{MN} = \mathbb{1}\Big(\max_{(m,n)\in\mathcal{K}_{M,N}} \tilde{t}^{\text{scan}}_{mn} > 1\Big),$$

with the statistic

$$\tilde{t}_{mn}^{\text{scan}} = \max_{\xi \in T_{M,N}(m,n)} \frac{1}{K\sqrt{mn}} \sum_{(i,j)\in A_\xi \times B_\xi} t_{ij,w^\star},$$

and $K^2 = K_{mn}^2 = 2(1+\delta)[(m \cdot \log(p^{-1}) + n \cdot \log(q^{-1})) + \log(MN)]$. The following theorem gives necessary conditions and a test procedure which adapts to the size (m,n) of the submatrix unknown within a given subset $\mathcal{K}_{M,N}$.

Theorem 4.1. *If $r_\epsilon \to 0$ and if, for each $(m,n) \in \mathcal{K}_{M,N}$, we have $a_{mn}(r_\epsilon)$ such that either*

$$\min_{(m,n)\in\mathcal{K}_{M,N}} a_{mn}^2(r_\epsilon) mnpq \to \infty$$

or

$$\liminf \min_{(m,n)\in\mathcal{K}_{M,N}} \frac{a_{mn}^2(r_\epsilon)}{2(m\cdot\log(p^{-1}) + n\cdot\log(q^{-1}))} > 1$$

then the test $\tilde{\psi} = \max\{\psi^{\chi^2}, \tilde{\psi}_{MN}^{\text{scan}}\}$, with $H = H_{MN} \to \infty$ such that $H_{MN} < c\min_{(m,n)\in\mathcal{K}_{M,N}} a_{m,n}(r_\epsilon)\sqrt{mnpq}$ for some $0 < c < 1$ and K_{mn} such that $\max_{(m,n)\in\mathcal{K}_{M,N}} K_{mn}^2 r_\epsilon^{1/\tau}/(mn) \to 0$, then $\gamma(\tilde{\psi}, \bigcup_{(m,n)\in\mathcal{K}_{M,N}} \Theta_{M,N}(\tau, r_\epsilon, m, n)) \to 0$.

Note that the separation rates are the same as in the nonadaptive case. Therefore, adaptive lower bounds are an immediate consequence of the non-adaptive lower bounds under the same additional conditions (and uniform in (m,n) over $\mathcal{K}_{M,N}$).

The test statistics t^{χ^2} and $\tilde{t}_{mn}^{\text{scan}}$ can be made free of r_ϵ by following the same procedure as in the previous section. They still achieve the same separation rates as in the case of known size (m,n).

4.4. *Implementation of the test procedures*

The linear procedure ψ^{χ^2} is rather simple to implement. However, there are difficulties for implementing the scan procedure ψ^{scan}. Indeed, computing the scan statistic t^{scan} implies computing standardized sums over all submatrices of size $m \times n$ in the large matrix $M \times N$. This is computationally infeasible for large values of M, N, m and n. However, a heuristic algorithm can be implemented as in Butucea and Ingster [4], following Sun and Nobel [21] and Shabalin *et al.* [22], which is a random procedure finding local maxima. It was observed by numerical simulations in [4] that with a sufficiently large choice of random initial values, the algorithm actually finds the global maximum that we aim at.

5. Proofs

Let us start with a preliminary result that gives an approximation of the moments generating function of $t_{ij,w^\star}$ defined in (2.1) under H_0.

Lemma 5.1. *For any real number λ such that $\lambda \sup_k w_k^\star = o(1)$,*

$$\mathbb{E}_0\bigl(\exp(\lambda t_{ij,w^\star})\bigr) = \exp\biggl(\frac{\lambda^2}{2}\bigl(1+o(1)\bigr)\biggr), \quad \forall (i,j) \in I \times J.$$

The proof of Lemma 5.1 is postponed in the Appendix.

5.1. *Proof of Theorem* 2.1

Observe that under H_0, $t_{ij,w}$ are i.i.d. random variables with zero mean and unit variance. Indeed, one gets $\mathrm{Var}_0(t_{ij,w}) = \sum_k w_k^2 \mathrm{Var}(\eta_{ij,k}^2)/\sigma_k^2 = 2\sum_k w_k^2 = 1$. Under the alternative, for all $\boldsymbol{\theta}_{ij} \in \Sigma(\tau, r_\epsilon)$,

$$\mathbb{E}_{\boldsymbol{\theta}_{ij}}(t_{ij,w}) = \sum_k w_k \frac{\theta_{ij,k}^2}{\sigma_k^2 \epsilon^2} \xi_{ij},$$

$$\begin{aligned}\mathrm{Var}_{\boldsymbol{\theta}_{ij}}(t_{ij,w}) &= \sum_k w_k^2 \left(2 + 4\frac{\theta_{ij,k}^2}{\sigma_k^2 \epsilon^2} \xi_{ij}\right) \\ &\le 1 + 4 \sup_k w_k \cdot \sum_k w_k \frac{\theta_{ij,k}^2}{\sigma_k^2 \epsilon^2} \xi_{ij} = 1 + 4 \sup_k w_k \cdot \mathbb{E}_{\boldsymbol{\theta}_{ij}}(t_{ij,w}).\end{aligned}$$

Due to the previous relations, for any $\overline{\boldsymbol{\theta}} \in \Theta_{M,N}(\tau, r_\epsilon, m, n)$

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\theta}}\bigl(t^{\chi^2}\bigr) &= \frac{1}{\sqrt{MN}} \sum_{(i,j)\in A_\xi \times B_\xi} \mathbb{E}_{\boldsymbol{\theta}_{ij}}(t_{ij,w^\star}) \\ &\ge \sqrt{MN} pq \cdot a(r_\epsilon) = \sqrt{mnpq} \cdot a(r_\epsilon),\end{aligned} \tag{5.1}$$

where the penultimate inequality follows from (2.5). Moreover, for the variance we have

$$\begin{aligned}\mathrm{Var}_{\boldsymbol{\theta}}\bigl(t^{\chi^2}\bigr) &= \frac{1}{MN} \sum_{(i,j)\in I \times J} \mathrm{Var}_{\boldsymbol{\theta}_{ij}}(t_{ij,w^\star}) \\ &= 1 + \frac{4}{MN} \sum_{(i,j)\in A_\xi \times B_\xi} \sum_k \bigl(w_k^\star\bigr)^2 \frac{\theta_{ij,k}^2}{\epsilon^2 \sigma_k^2} \\ &\le 1 + 4 \sup_k w_k^\star \frac{1}{\sqrt{MN}} \mathbb{E}_{\boldsymbol{\theta}}\bigl(t^{\chi^2}\bigr).\end{aligned} \tag{5.2}$$

Recall that $\sup_k w_k^\star \overset{r_\epsilon \to 0}{\longrightarrow} 0$ (see Proposition 2.1).

Type I error probability of ψ^{χ^2}. Since $\sup_k w_k^\star = o(1)$, the asymptotic standard normality of t^{χ^2} under the null follows from Lemma 3.1 in Ingster and Suslina [17] then, as H large enough,

$$\mathbb{P}_0\bigl(t^{\chi^2} > H\bigr) = \Phi(-H) + o(1),$$

where Φ stands for the c.d.f. of a standard Gaussian random variable.

Maximal type II error probability of ψ^{χ^2} *uniformly over* $\Theta_{M,N}(\tau, r_\epsilon, m, n)$. We deduce from (5.2) that $\mathrm{Var}_{\overline{\boldsymbol{\theta}}}(t^{\chi^2}) = 1 + o(\mathbb{E}_{\overline{\boldsymbol{\theta}}}(t^{\chi^2}))$, uniformly over $\overline{\boldsymbol{\theta}} \in \Theta_{M,N}(\tau, r_\epsilon, n, m)$.

For all $\overline{\boldsymbol{\theta}}$ in $\Theta_{M,N}(\tau, r_\epsilon, m, n)$, by Markov's inequality and relation (5.1),

$$\begin{aligned}\mathbb{P}_{\overline{\boldsymbol{\theta}}}\bigl(t^{\chi^2} \le H\bigr) &\le \mathbb{P}_{\overline{\boldsymbol{\theta}}}\bigl(\bigl|t^{\chi^2} - \mathbb{E}_{\overline{\boldsymbol{\theta}}}\bigl(t^{\chi^2}\bigr)\bigr| \ge \mathbb{E}_{\overline{\boldsymbol{\theta}}}\bigl(t^{\chi^2}\bigr) - H\bigr) \\ &\le \frac{\mathrm{Var}_{\overline{\boldsymbol{\theta}}}(t^{\chi^2})}{(\mathbb{E}_{\overline{\boldsymbol{\theta}}}(t^{\chi^2}) - H)^2} \\ &\le \frac{1 + 4\sup_k w_k^\star \mathbb{E}_{\overline{\boldsymbol{\theta}}}(t^{\chi^2})/\sqrt{MN}}{(\mathbb{E}_{\overline{\boldsymbol{\theta}}}(t^{\chi^2}) - H)^2}\end{aligned}$$

$$\leq \frac{1}{(1-c)^2 \mathbb{E}_{\overline{\boldsymbol{\theta}}}(t^{\chi^2})^2} + \frac{4 \sup_k w_k^\star}{(1-c)^2 \sqrt{MN} \mathbb{E}_{\overline{\boldsymbol{\theta}}}(t^{\chi^2})}$$
$$= o(1),$$

provided that $a(r_\epsilon)\sqrt{mnpq} \to +\infty$ and $H \leq c \cdot a(r_\epsilon)\sqrt{mnpq}$ for some $0 < c < 1$.

Type I error probability of ψ^{scan}.

Applying Markov's inequality,

$$\begin{aligned}
\mathbb{P}_0\big(t^{\text{scan}} > K\big) &\leq \sum_{\xi \in T_{M,N}(m,n)} \mathbb{P}_0\bigg(\frac{1}{\sqrt{mn}} \sum_{(i,j)\in A_\xi \times B_\xi} t_{ij,w^\star} > K\bigg) \\
&= \binom{M}{m}\binom{N}{n} \mathbb{P}_0\bigg(\frac{1}{\sqrt{mn}} \sum_{(i,j)\in A_\xi \times B_\xi} t_{ij,w^\star} > K\bigg) \\
&\leq \binom{M}{m}\binom{N}{n} \exp\big(-K^2\big)\mathbb{E}_0\bigg(\exp\bigg(\sum_{(i,j)\in A_\xi \times B_\xi} \frac{K}{\sqrt{mn}} t_{ij,w^\star}\bigg)\bigg) \\
&\leq \binom{M}{m}\binom{N}{n} \exp\big(-K^2\big)\bigg(\mathbb{E}_0\bigg(\exp\bigg(\frac{K}{\sqrt{mn}} t_{11,w^\star}\bigg)\bigg)\bigg)^{mn}.
\end{aligned} \tag{5.3}$$

Set $\lambda = K/\sqrt{mn}$ with $K = \sqrt{2(1+\delta)\log(\binom{N}{n}\binom{M}{m})}$, for some small $\delta > 0$ and note that $K/\sqrt{mn}\sup_k w_k^\star \leq K r_\epsilon^{1/(2\tau)}/\sqrt{mn} = o(1)$ by assumption in our theorem; then, applying Lemma 5.1 we obtain that

$$\mathbb{E}_0\big(\exp(\lambda t_{11,w^\star})\big) = \exp\bigg(\frac{\lambda^2}{2}\big(1+o(1)\big)\bigg).$$

Next, by plugging $(\mathbb{E}_0(\exp(\frac{K}{\sqrt{nm}} t_{11,w^\star})))^{nm} = \exp(\frac{K^2}{2}(1+o(1)))$ into (5.3), we obtain

$$\mathbb{P}_0\big(t^{\text{scan}} > K\big) \leq \binom{M}{m}\binom{N}{n} \exp\big(-K^2/2\big(1+o(1)\big)\big) = o(1), \tag{5.4}$$

due to the choice of $K = \sqrt{2(1+\delta)\log(\binom{N}{n}\binom{M}{m})}$, for some small $\delta > 0$.

Maximal Type II error probability of ψ^{scan} *uniformly over* $\Theta_{M,N}(\tau, r_\epsilon, m, n)$. For any $\overline{\boldsymbol{\theta}} \in \Theta_{M,N}(\tau, r_\epsilon, m, n)$, it exists $A \subset I$ and $B \subset J$ such that $\#A = m$, $\#B = n$ and $\xi_{ij} = \mathbb{1}((i,j) \in A \times B)$; using the inequality $t^{\text{scan}} \geq \frac{1}{\sqrt{mn}} \sum_{(i,j)\in A\times B} t_{ij,w^\star}$, we obtain

$$\begin{aligned}
\mathbb{P}_{\overline{\boldsymbol{\theta}}}\big(t^{\text{scan}} \leq K\big) &\leq \mathbb{P}_{\overline{\boldsymbol{\theta}}}\bigg(\frac{1}{\sqrt{mn}} \sum_{(i,j)\in A\times B} t_{ij,w^\star} \leq K\bigg) \\
&\leq \frac{\operatorname{Var}_{\overline{\boldsymbol{\theta}}}(1/(\sqrt{mn}) \sum_{(i,j)\in A\times B} t_{ij,w^\star})}{(\mathbb{E}_{\overline{\boldsymbol{\theta}}}(1/(\sqrt{mn}) \sum_{(i,j)\in A\times B} t_{ij,w^\star}) - K)^2}.
\end{aligned}$$

Due to (2.5), we have

$$\mathbb{E}_{\overline{\boldsymbol{\theta}}}\bigg(\frac{1}{\sqrt{mn}} \sum_{(i,j)\in A\times B} t_{ij,w^\star}\bigg) = \frac{1}{\sqrt{mn}} \sum_{(i,j)\in A\times B} \mathbb{E}_{\boldsymbol{\theta}_{ij}}(t_{ij,w^\star}) \geq a(r_\epsilon)\sqrt{mn}.$$

By assumption (2.9) we have $\liminf a(r_\epsilon)\sqrt{mn}/K \geq (1+\delta)^{-1/2}$, which implies that, asymptotically, $K \leq a(r_\epsilon)\sqrt{mn(1+\delta)}/(1+\tilde{\delta})$ for some $\tilde{\delta} > 0$ and then $K \leq c a(r_\epsilon)\sqrt{mn} \leq c\mathbb{E}_{\overline{\boldsymbol{\theta}}}(\frac{1}{\sqrt{mn}} \sum_{(i,j)\in A\times B} t_{ij,w^\star})$ for some $0 < c < 1$ if δ is small enough.

Now, acting as for getting Equation (5.2), we have

$$\begin{aligned}\mathrm{Var}_{\overline{\boldsymbol{\theta}}}\biggl(\frac{1}{\sqrt{mn}}\sum_{(i,j)\in A\times B} t_{ij,w^\star}\biggr) &= \frac{1}{mn}\sum_{(i,j)\in A\times B}\mathrm{Var}_{\boldsymbol{\theta}_{ij}}(t_{ij,w^\star})\\ &\le 1+\frac{4\sup_k w_k^\star}{mn}\sum_{(i,j)\in A\times B}\mathbb{E}_{\boldsymbol{\theta}_{ij}}(t_{ij,w^\star})\\ &\le 1+\frac{4\sup_k w_k^\star}{\sqrt{mn}}\mathbb{E}_{\overline{\boldsymbol{\theta}}}\biggl(\frac{1}{\sqrt{mn}}\sum_{(i,j)\in A\times B} t_{ij,w^\star}\biggr).\end{aligned}$$

Finally,

$$\mathbb{P}_{\overline{\boldsymbol{\theta}}}\bigl(t^{\mathrm{scan}}\le K\bigr)\le\frac{1}{(1-c)^2a^2(r_\epsilon)mn}+\frac{4\sup_k w_k^\star}{(1-c)^2a(r_\epsilon)mn}=o(1).$$

5.2. *Proof of Theorem* 4.1

As the test procedure ψ^{χ^2} in (2.6) does not depend on (m,n) the same upper bounds hold for it. The proof is slightly different for $\tilde{\psi}^{\mathrm{scan}}_{MN}$.

The type I error probability is bounded from above as in (5.4) for the modified value of K in this theorem:

$$\begin{aligned}\mathbb{P}_0\Bigl(\max_{(m,n)\in\mathcal{K}_{M,N}}\tilde{t}^{\mathrm{scan}}_{mn}>1\Bigr) &\le \sum_{(m,n)\in\mathcal{K}_{M,N}}\mathbb{P}_0\bigl(t^{\mathrm{scan}}>K\bigr)\\ &\le \sum_{(m,n)\in\mathcal{K}_{M,N}}(MN)^{-(1+\delta)}\exp\biggl(-\delta\log\binom{M}{m}\binom{N}{n}\bigl(1+o(1)\bigr)\biggr),\end{aligned}$$

for $0<\delta<1$ small. By hypothesis, $\delta\log\bigl(\binom{M}{m}\bigr)\asymp\delta m\log(p^{-1})\gg\Delta m$, for large m,M and some $\Delta>1$, as $p\to 0$ uniformly. Therefore,

$$\mathbb{P}_0\Bigl(\max_{(m,n)\in\mathcal{K}_{M,N}}\tilde{t}^{\mathrm{scan}}_{mn}>1\Bigr)\le\sum_{(m,n)\in\mathcal{K}_{M,N}}(\Delta)^{-mn}(MN)^{-(1+\delta)}=o(1).$$

As for the type II error probability, we first fix (m,n) in $\mathcal{K}_{M,N}$, then fix a matrix ξ in $T_{M,N}$ and then the same proof in Theorem 2.1 holds.

5.3. *Proof of Theorem* 3.1

The usual steps for proving the lower bounds are the following. First, we bound from below the minimax total error probability by reducing the set of parameters. Next, we choose a prior on the reduced set of parameters and bound the testing risk from below with a Bayesian risk. Finally, this Bayesian risk is large if a χ^2-distance between the likelihoods under the null and under the mixture of alternatives is small.

We follow closely the proof in Butucea and Ingster [4] with important modifications; indeed, the two-sided alternative involves a Bayesian prior on the sequences $\{\pm\theta_k^*\}_k$ and for the study of the averaged log-likelihood with respect to this prior, instead of the Laplace transform of a Gaussian, we give in Lemma A.1 the asymptotic behaviour of the Laplace transform of a random series. The latter is approximately the same as in the Gaussian case in our restricted range of parameters.

Recall that $\{\theta_k^\star\}_{k\in\mathbb{Z}}$ is the solution of the optimisation problem (2.2) and let us choose a matrix ξ in the set $T_{M,N}(m,n)$, $\xi=\mathbb{1}((i,j)\in A\times B)$ where $A=A_\xi$ is a set of m rows out of M and $B=B_\xi$ a set of n columns out of N. Denote by

$$\mathcal{T}_{M,N}(\tau,r_\epsilon,m,n)=\bigl\{\overline{\boldsymbol{\theta}}=\bigl[\xi_{ij}\bigl\{\pm\theta_k^\star\bigr\}_k\bigr]_{(i,j)\in I\times J},\xi\in T_{M,N}(m,n)\bigr\}.$$

This is the reduced set of parameters, i.e., a subset of the alternative $\Theta_{M,N}(\tau,r_\epsilon,m,n)$ in our test.

A prior measure on the reduced set will choose ξ with equal probability in the set $T_{M,N}(m,n)$; given ξ, the (θ_{ij})'s associated with non-zero ξ_{ij} are i.i.d. and for (i,j) such that $\xi_{ij}=1$, the prior will choose with equal probability between $\theta_k^\star$ and $-\theta_k^\star$, independently for each k. We can write $\pi_{ij,k}=\frac{1}{2}(\delta_{-\theta_k^\star}+\delta_{\theta_k^\star})$, where δ stands for the Dirac measure, and $\pi_{ij}=\prod_k \pi_{ij,k}$. Let us define

$$\overline{\pi}=\frac{1}{\binom{M}{m}\binom{N}{n}}\sum_{\xi\in T_{M,N}(m,n)}\prod_{(i,j)\in A_\xi\times B_\xi}\pi_{ij}$$

the global prior on $\overline{\boldsymbol{\theta}}$'s in $\mathcal{T}_{M,N}(\tau,r_\epsilon,m,n)$.

Let us write the likelihood ratio of one active component, i.e., when (i,j) is such that $\xi_{ij}=1$,

$$\frac{d\mathbb{P}_{\pi_{ij}}}{d\mathbb{P}_0}\big(\{x_{ij,k}\}_k\big)=\prod_k \exp\biggl(-\frac{\theta_k^{\star 2}}{2\epsilon^2\sigma_k^2}\biggr)\cosh\biggl(x_{ij,k}\frac{\theta_k^\star}{\epsilon^2\sigma_k^2}\biggr). \tag{5.5}$$

Set $\overline{X}=[\{x_{ij,k}\}_k]_{(i,j)}$. Then the likelihood ratio with respect to the null hypothesis of our observations becomes:

$$L_{\overline{\pi}}(\overline{X})=\frac{d\mathbb{P}_{\overline{\pi}}}{d\mathbb{P}_0}\big(\big[\{x_{ij,k}\}_k\big]_{(i,j)}\big)=\frac{1}{\binom{M}{m}\binom{N}{n}}\sum_{\xi\in T_{M,N}(m,n)}\frac{d\mathbb{P}_\xi}{d\mathbb{P}_0}(\overline{X}),$$

where

$$\frac{d\mathbb{P}_\xi}{d\mathbb{P}_0}(\overline{X})=\prod_{(i,j)\in A_\xi\times B_\xi}\frac{d\mathbb{P}_{\pi_{ij}}}{d\mathbb{P}_0}\big(\{x_{ij,k}\}_k\big). \tag{5.6}$$

In order to prove that we cannot distinguish the hypotheses asymptotically, we see that

$$\begin{aligned}
\gamma &= \inf_{\psi\in[0,1]}\Bigl(w(\psi)+\sup_{\overline{\boldsymbol{\theta}}\in\Theta_{M,N}(\tau,r_\epsilon,m,n)}\mathbb{E}_{\overline{\boldsymbol{\theta}}}\bigl[1-\psi(\overline{X})\bigr]\Bigr)\\
&\ge \inf_{\psi\in[0,1]}\Bigl(w(\psi)+\sup_{\overline{\boldsymbol{\theta}}\in\mathcal{T}_{M,N}(\tau,r_\epsilon,m,n)}\mathbb{E}_{\overline{\boldsymbol{\theta}}}\bigl[1-\psi(\overline{X})\bigr]\Bigr)\\
&\ge \inf_{\psi\in[0,1]}\Biggl(w(\psi)+\sum_{\overline{\boldsymbol{\theta}}\in\mathcal{T}_{M,N}(\tau,r_\epsilon,m,n)}\overline{\pi}(\overline{\boldsymbol{\theta}})\mathbb{E}_{\overline{\boldsymbol{\theta}}}\bigl[1-\psi(\overline{X})\bigr]\Biggr)\\
&\ge \inf_{\psi\in[0,1]}\bigl(\mathbb{E}_0\bigl(\psi(\overline{X})\bigr)+\mathbb{E}_0\bigl[\bigl(1-\psi(\overline{X})\bigr)L_{\overline{\pi}}(\overline{X})\bigr]\bigr).
\end{aligned}$$

This infimum is attained for the likelihood ratio test $\psi^\star(\overline{X})=\mathbb{1}(L_{\overline{\pi}}(\overline{X})>1)$. By Fatou lemma, we have

$$\liminf\gamma\ge\mathbb{E}_0\bigl(\liminf\bigl(\psi^\star(\overline{X})+\bigl(1-\psi^\star(\overline{X})\bigr)L_{\overline{\pi}}(\overline{X})\bigr)\bigr),$$

which implies that $\gamma\to 1$ if $L_{\overline{\pi}}(\overline{X})\to 1$ in $\mathbb{P}_0$-probability. In order to prove this sufficient condition, it is enough to check that

$$\mathbb{E}_0\bigl(L_{\overline{\pi}}(\overline{X})^2\bigr)\le 1+o(1). \tag{5.7}$$

First, let us consider $\mathbb{E}_0(L_{\overline{\pi}}^2(\overline{X}))$ and let us see that (5.7) can not be obtained; the explanation is that too many events with small probability are summed up in the expected value of the square likelihood ratio.

Non-truncated likelihood. Let us denote by $\mathcal{X}$ the random matrix uniformly distributed on $T_{M,N}(m,n)$, i.e. $P(\mathcal{X}=\xi)=(\binom{M}{m}\binom{N}{n})^{-1}$. Then, using (5.5) and (5.6), one has

$$\begin{aligned}\mathbb{E}_0\big(L_{\pi}^2(\overline{X})\big) &= \frac{1}{(\binom{M}{m}\binom{N}{n})^2}\mathbb{E}_0\bigg(\sum_{\xi_1}\sum_{\xi_2}\frac{d\mathbb{P}_{\xi_1}}{d\mathbb{P}_0}(\overline{X})\frac{d\mathbb{P}_{\xi_2}}{d\mathbb{P}_0}(\overline{X})\bigg) \\ &= \frac{1}{(\binom{M}{m}\binom{N}{n})^2}\sum_{\xi_1}\sum_{\xi_2}\mathbb{E}_0\bigg(\prod_{(i,j)\in A_{\xi_1}\times B_{\xi_1}}\prod_k \exp\bigg(-\frac{\theta_k^{\star 2}}{2\epsilon^2\sigma_k^2}\bigg)\cosh\bigg(x_{ij,k}\frac{\theta_k^\star}{\epsilon^2\sigma_k^2}\bigg) \\ &\quad\times\prod_{(u,v)\in A_{\xi_2}\times B_{\xi_2}}\prod_k \exp\bigg(-\frac{\theta_k^{\star 2}}{2\epsilon^2\sigma_k^2}\bigg)\cosh\bigg(x_{uv,k}\frac{\theta_k^\star}{\epsilon^2\sigma_k^2}\bigg)\bigg).\end{aligned}$$

Using Laplace transform of standard Gaussian random variable, note that, if $Z\sim\mathcal{N}(0,1)$ and if $\lambda>0$

$$\mathbb{E}\big(\cosh(Z\lambda)\exp(-\lambda^2/2)\big)=1 \quad\text{and}\quad \mathbb{E}\big(\cosh^2(Z\lambda)\exp(-\lambda^2)\big)=\cosh(\lambda^2),$$

that we apply to $x_{ij,k}/(\epsilon\sigma_k)$. We deduce that

$$\begin{aligned}\mathbb{E}_0\big(L_{\pi}^2(\overline{X})\big) &= \frac{1}{(\binom{M}{m}\binom{N}{n})^2}\sum_{\xi_1}\sum_{\xi_2}\prod_{(i,j)\in A_{\xi_1}\cap A_{\xi_2}\times B_{\xi_1}\cap B_{\xi_2}}\prod_k\cosh\bigg(\frac{\theta_k^{\star 2}}{\epsilon^2\sigma_k^2}\bigg) \\ &= \mathbb{E}_{\mathcal{X}_1,\mathcal{X}_2}\exp(\bigg(H(\mathcal{X}_1,\mathcal{X}_2)L(\mathcal{X}_1,\mathcal{X}_2)\log\bigg(\prod_k\cosh\bigg(\frac{\theta_k^{\star 2}}{\epsilon^2\sigma_k^2}\bigg)\bigg)\bigg),\end{aligned}$$

where $H(\mathcal{X}_1,\mathcal{X}_2)$ and $L(\mathcal{X}_1,\mathcal{X}_2)$ are random variables such that $H(\xi_1,\xi_2)=\#A_{\xi_1}\cap A_{\xi_2}$ and $L(\xi_1,\xi_2)=\#B_{\xi_1}\cap B_{\xi_2}$ with ξ_1,ξ_2 realizations of $\mathcal{X}_1,\mathcal{X}_2$.

At this stage, let us evaluate $D:=\log(\prod_k\cosh(\frac{\theta_k^{\star 2}}{\epsilon^2\sigma_k^2}))$:

$$\begin{aligned}D &= \sum_k\log\bigg(1+2\sinh^2\bigg(\frac{\theta_k^{\star 2}}{2\epsilon^2\sigma_k^2}\bigg)\bigg)=\sum_k\log\bigg(1+2\bigg(\frac{\theta_k^{\star 2}}{2\epsilon^2\sigma_k^2}\bigg)^2\big(1+o(1)\big)\bigg) \\ &= \sum_k\bigg(\frac{\theta_k^{\star 4}}{2\epsilon^4\sigma_k^4}\bigg)\big(1+o(1)\big)=a^2(r_\epsilon)\big(1+o(1)\big),\end{aligned} \tag{5.8}$$

which holds under relation (2.4).

Back to $\mathbb{E}_0(L_{\pi}^2(\overline{X}))$, we can write

$$\mathbb{E}_0\big(L_{\pi}^2(\overline{X})\big)=\mathbb{E}_{\mathcal{X}_1,\mathcal{X}_2}\big(\mathbb{E}_{\mathcal{X}_1,\mathcal{X}_2}^{\mathcal{X}_1}\big(\exp\big(H(\mathcal{X}_1,\mathcal{X}_2)L(\mathcal{X}_1,\mathcal{X}_2)D\big)\big)\big).$$

Note that, given $\mathcal{X}_1$, $H(\mathcal{X}_1,\mathcal{X}_2)\sim\mathcal{HG}(M,m,m)$, $L(\mathcal{X}_1,\mathcal{X}_2)\sim\mathcal{HG}(N,n,n)$ and $H(\mathcal{X}_1,\mathcal{X}_2)$ and $L(\mathcal{X}_1,\mathcal{X}_2)$ are independent random variables. Then, due to Proposition 20.6, page 173 in [1], there exist two Binomial variables $Y_1\sim B(m,p)$ and $Y_2\sim B(n,q)$ such that $\mathbb{E}_{\mathcal{X}_1,\mathcal{X}_2}^{\mathcal{X}_1}(Y_1|H(\mathcal{X}_1,\mathcal{X}_2))=H(\mathcal{X}_1,\mathcal{X}_2)$ and $\mathbb{E}_{\mathcal{X}_1,\mathcal{X}_2}^{\mathcal{X}_1}(Y_2|L(\mathcal{X}_1,\mathcal{X}_2))=L(\mathcal{X}_1,\mathcal{X}_2)$, where $\mathbb{E}_{\mathcal{X}_1,\mathcal{X}_2}^{\mathcal{X}_1}$ denotes that conditional expectation w.r.t. $\mathcal{X}_1$.

Now, applying twice Jensen's Inequality and then using the generating function of Binomial random variables, we deduce

$$\begin{aligned}\mathbb{E}_0\big(L_{\pi}^2(\overline{X})\big) &= \mathbb{E}_{\mathcal{X}_1,\mathcal{X}_2}\big(\mathbb{E}_{\mathcal{X}_1,\mathcal{X}_2}^{\mathcal{X}_1}\big(\exp\big(\mathbb{E}_{\mathcal{X}_1,\mathcal{X}_2}^{\mathcal{X}_1}\big(Y_1|H(\mathcal{X}_1,\mathcal{X}_2)\big)\mathbb{E}_{\mathcal{X}_1,\mathcal{X}_2}^{\mathcal{X}_1}\big(Y_2|L(\mathcal{X}_1,\mathcal{X}_2)\big)D\big)\big)\big) \\ &\le \mathbb{E}_{\mathcal{X}_1,\mathcal{X}_2}\mathbb{E}_{\mathcal{X}_1,\mathcal{X}_2}^{\mathcal{X}_1}\big(\exp\big(Y_1\mathbb{E}_{\mathcal{X}_1,\mathcal{X}_2}^{\mathcal{X}_1}\big(Y_2|L(\mathcal{X}_1,\mathcal{X}_2)\big)D\big)|H(\mathcal{X}_1,\mathcal{X}_2)\big)\end{aligned}$$

$$
\begin{aligned}
&\le \mathbb{E}\big(\exp(Y_1Y_2D)\big)\\
&= \mathbb{E}\mathbb{E}\big(\exp(Y_1Y_2D)|Y_1\big)\\
&= \mathbb{E}\big(\big[\big(\exp(Y_1D)-1\big)q+1\big]^n\big)\\
&= \mathbb{E}\big(\exp\big(n\log\big(1+\big(\exp(Y_1D)-1\big)q\big)\big)\big).
\end{aligned}
\tag{5.9}
$$

Next by applying twice the Mean Value Theorem, using twice the inequality $\log(1+x)\le x$ on $\mathbf{R}^+$ and again due to generating functions of Binomial random variables, we obtain that there exist $\theta_1\in]0,1[$ and $\theta_2\in]0,1[$

$$
\begin{aligned}
\mathbb{E}_0\big(L^2_{\overline{\pi}}(\overline{X})\big) &\le \mathbb{E}\big(\exp\big(n\log\big(1+Y_1qD\exp(\theta_1Y_1D)\big)\big)\big)\\
&\le \mathbb{E}\big(\exp\big(nY_1qD\exp(\theta_1mD)\big)\big)\\
&= \big[p\big(\exp\big(nqD\exp(\theta_1mD)\big)-1\big)+1\big]^m\\
&= \exp\big(m\log\big(1+pnqD\exp(\theta_1mD)\exp\big(\theta_2nqD\exp(\theta_1mD)\big)\big)\big)\\
&\le \exp\big(mpnqD\exp(\theta_1mD)\exp\big(\theta_2nqD\exp(\theta_1mD)\big)\big),
\end{aligned}
$$

in which the argument of the exponential can not be made arbitrary small.

Truncated likelihood. That is why we have to modify slightly the likelihood ratio, by truncation, as follows:

$$
\tilde{L}_{\overline{\pi}}(\overline{X}) = \frac{1}{\binom{M}{m}\binom{N}{n}}\sum_{\xi\in T_{M,N}(m,n)}\prod_{(i,j)\in A_\xi\times B_\xi}\prod_k \exp\left(-\frac{\theta_k^{\star 2}}{2\epsilon^2\sigma_k^2}\right)\cosh\left(x_{ij,k}\frac{\theta_k^\star}{\epsilon^2\sigma_k^2}\right)\mathbb{1}(\Gamma_\xi),
$$

where the event Γ_ξ is defined later on.

The idea is that the random variable in this event is truncated at the values predicted by large deviations and this is sufficient to diminish the second-order moment of the likelihood ratio.

Let us denote $\Gamma=\bigcap_{\xi\in T_{M,N}(m,n)}\Gamma_\xi$ and Γ^C its complement. Then, for some fixed $\delta>0$, let us consider the event $\mathcal{E}=\{|L_{\overline{\pi}}(\overline{X})-1|>\delta\}$

$$
\begin{aligned}
\mathbb{P}_0(\mathcal{E}) &= \mathbb{P}_0(\mathcal{E}\cap\Gamma)+\mathbb{P}_0\big(\mathcal{E}\cap\Gamma^C\big)\\
&\le \delta^{-2}\mathbb{E}_0\big(\big(L_{\overline{\pi}}(\overline{X})-\tilde{L}_{\overline{\pi}}(\overline{X})+\tilde{L}_{\overline{\pi}}(\overline{X})-1\big)^2\mathbb{1}(\Gamma)\big)+\mathbb{P}_0\big(\Gamma^C\big)\\
&\le 2\delta^{-2}\big\{\mathbb{E}_0\big(\big(L_{\overline{\pi}}(\overline{X})-\tilde{L}_{\overline{\pi}}(\overline{X})\big)^2\mathbb{1}(\Gamma)\big)+\mathbb{E}_0\big(\big(\tilde{L}_{\overline{\pi}}(\overline{X})-1\big)^2\mathbb{1}(\Gamma)\big)\big\}+\mathbb{P}_0\big(\Gamma^C\big)\\
&\le 2\delta^{-2}\mathbb{E}_0\big(\big(\tilde{L}_{\overline{\pi}}(\overline{X})-1\big)^2\big)+\mathbb{P}_0\big(\Gamma^C\big)\\
&\le 2\delta^{-2}\big[\big(\mathbb{E}_0\big(\tilde{L}_{\overline{\pi}}(\overline{X})^2\big)-1\big)-2\big(\mathbb{E}_0\big(\tilde{L}_{\overline{\pi}}(\overline{X})\big)-1\big)\big]+\mathbb{P}_0\big(\Gamma^C\big),
\end{aligned}
$$

where we used the following equality $\mathbb{1}(\Gamma)L_{\overline{\pi}}(\overline{X})=\mathbb{1}(\Gamma)\tilde{L}_{\overline{\pi}}(\overline{X})$. Then, it remains to prove the following lemma to complete the proof of Theorem 3.1.

Lemma 5.2. *Under the assumptions of Theorem 3.1 we have the following*:

1. $\mathbb{P}_0(\Gamma)\to 1$.
2. $\mathbb{E}_0(\tilde{L}_{\overline{\pi}}(\overline{X}))\to 1$.
3. $\mathbb{E}_0(\tilde{L}_{\overline{\pi}}(\overline{X})^2)\le 1+o(1)$.

The proof of Lemma 5.2 is postponed in Section A.1.

In order to finish this part, let us define the event Γ_ξ for some small $\delta_1>0$ as follows

$$
\Gamma_\xi = \bigcap_{\substack{\delta_1 m\le h\le m,\\ \delta_1 n\le l\le n}}\ \bigcap_{\substack{V\in T_{M,N}(h,l):\\ A_V\subset A_\xi, B_V\subset B_\xi}}\{Y_V\le T_{hl}\},
\tag{5.10}
$$

where

$$\begin{cases} Y_V = \frac{1}{a(r_\epsilon)\sqrt{hl}} \sum_{(i,j)\in A_V\times B_V} \sum_k (\log\cosh(u_k \cdot \frac{x_{ij,k}}{\epsilon\sigma_k}) - \frac{u_k^2}{2} + \frac{u_k^4}{4}), \quad u_k = \theta_k^\star/(\epsilon\sigma_k), \\ T_{hl}^2 = 2\big(\log(\binom{M}{h}\binom{N}{l})(1+A_\epsilon) + \log(mn)\big), \end{cases}$$

for

$$A_\epsilon = \max\left\{\frac{1}{\log\log(\binom{M}{h}\binom{N}{l})}, \frac{1}{\log((a(r_\epsilon)\sup_k w_k^\star)^{-1})}\right\}.$$

Note that $\theta_k^\star$ is null for $k > T$ and thus the sum over k in Y_V contains a finite number of non-null elements. Due to (2.4), recall that we have

$$\sum_k u_k^4 = 2a^2(r_\epsilon). \tag{5.11}$$

We want T_{hl}^2 to be equivalent to $2\big(\log(\binom{M}{h}\binom{N}{l}) + \log(mn)\big)$, therefore we should have $A_\epsilon = o(1)$. Moreover, the proof of the first item in Lemma 5.2, requires that

$$A_\epsilon \cdot \log\left(\binom{M}{h}\binom{N}{l}\right) \to \infty \quad \text{and} \quad A_\epsilon\Big/\Big(a(r_\epsilon)\sup_k w_k^\star\Big) \to \infty. \tag{5.12}$$

Let us see that with our choice of A_ϵ, the relations (5.12) as well as $A_\epsilon = o(1)$ are satisfied for any r_ϵ such that $r_\epsilon < \tilde{r}_\epsilon$,

$$\begin{aligned} a(r_\epsilon)\cdot\sup_k w_k^\star &\le O(1)a(\tilde{r}_\epsilon)\tilde{r}_\epsilon^{1/(2\tau)} \\ &\le O(1)a(\tilde{r}_\epsilon)^{1+1/(4\tau+4s+1)}\epsilon^{2/(4\tau+4s+1)} \\ &\le O(1)\big(a(\tilde{r}_\epsilon)\cdot\epsilon^{1/(2\tau+2s+1)}\big)^{(4\tau+4s+2)/(4\tau+4s+1)} = o(1), \end{aligned}$$

by assumption (3.3). This also implies that $A_\epsilon = o(1)$. Moreover,

$$A_\epsilon\Big/\Big(a(r_\epsilon)\sup_k w_k^\star\Big) \ge \left(\log\left(\Big(a(r_\epsilon)\sup_k w_k^\star\Big)^{-1}\right)\right)^{-1}\cdot\Big(a(r_\epsilon)\sup_k w_k^\star\Big)^{-1} \to \infty.$$

Also, $A_\epsilon\cdot\log(\binom{M}{h}\binom{N}{l}) \ge \log(\binom{M}{h}\binom{N}{l})/\log\log(\binom{M}{h}\binom{N}{l}) \to \infty$.

Appendix

Lemma A.1. *If $r_\epsilon \to 0$ such that $a(r_\epsilon)\cdot\sup_k w_k^\star = o(1)$, then for any $\lambda > 0$ such that $\lambda = O(1)$,*

$$\mathbb{E}_0(\exp\left(\lambda\sum_k\left(\log\cosh(u_k\cdot\eta_{11,k}) - \frac{u_k^2}{2} + \frac{u_k^4}{4}\right)\right) = \exp\left(\frac{\lambda^2 a^2(r_\epsilon)}{2}\left(1 + O\Big(a(r_\epsilon)\sup_k w_k^\star\Big)\right)\right).$$

Proof. Let us see that for bounded $\lambda > 0$, for $u \to 0$ and a standard Gaussian random variable η, we have:

$$\mathbb{E}\left(\exp\left(\lambda\cdot\left(\log\cosh(u\cdot\eta) - \frac{u^2}{2} + \frac{u^4}{4} + O\big(u^6\big)\right)\right)\right) = \exp\left(\frac{\lambda^2 u^4}{4} + O\big(u^6\big)\right).$$

This proof can be adapted from Gayraud and Ingster [7] (cf. Lemma A.1). Now, we apply this result for each k and recall that $u_k^2 = a(r_\epsilon) \cdot w_k^\star \le a(r_\epsilon) \cdot \sup_k w_k^\star = o(1)$ by assumption. Using $\sum_k u_k^4 = 2a^2(r_\epsilon)$, we get

$$\mathbb{E}_0\left(\exp\left(\lambda \sum_k \left(\log\cosh(u_k \cdot \eta_{11,k}) - \frac{u_k^2}{2} + \frac{u_k^4}{4}\right)\right)\right) = \exp\left(\frac{\lambda^2}{4} \sum_k u_k^4\left(1 + O\left(u_k^2\right)\right)\right)$$
$$= \exp\left(\frac{\lambda^2 a^2(r_\epsilon)}{2}\left(1 + O\left(a(r_\epsilon) \cdot \sup_k w_k^\star\right)\right)\right). \qquad \square$$

A.1. *Proof of Lemma* 5.2

Take a small $\delta > 0$. The detection boundary $a(r_\epsilon)$ satisfies (3.5), so the most difficult case is when the limit is close to 1. Therefore, we shall assume that

$$a^2(r_\epsilon)mn \sim (2-\delta)\left(m \cdot \log\left(p^{-1}\right) + n \cdot \log\left(q^{-1}\right)\right).$$

This implies

$$a^2(r_\epsilon) \asymp \frac{\log(p^{-1})}{n} + \frac{\log(q^{-1})}{m}. \tag{A.1}$$

1. We shall prove that $\mathbb{P}_0(\Gamma^C) \to 0$. Let us write more conveniently

$$\Gamma^C = \bigcup_{\xi \in T_{M,N}(m,n)} \bigcup_{\substack{\delta_1 m \le h \le m,\\ \delta_1 n \le l \le n}} \bigcup_{\substack{V \subset \xi:\\ A_V \subset A_\xi, B_V \subset B_\xi}} \{Y_V > T_{hl}\}$$
$$= \bigcup_{\substack{\delta_1 m \le h \le m,\\ \delta_1 n \le l \le n}} \bigcup_{V \in T_{M,N}(h,l)} \{Y_V > T_{hl}\}.$$

Therefore, we have

$$\mathbb{P}_0\left(\Gamma^C\right) \le \sum_{\substack{\delta_1 m \le h \le m,\\ \delta_1 n \le l \le n}} \sum_{V \in T_{M,N}(h,l)} \mathbb{P}_0\left(Y_V T_{hl} > T_{hl}^2\right)$$
$$\le \sum_{\substack{\delta_1 m \le h \le m,\\ \delta_1 n \le l \le n}} \sum_{V \in T_{M,N}(h,l)} \exp\left(-T_{hl}^2\right)\mathbb{E}_0\left(\exp\left(\sum_{(i,j) \in A_V \times B_V} T_{hl} Y_V\right)\right)$$
$$\le \sum_{\substack{\delta_1 m \le h \le m,\\ \delta_1 n \le l \le n}} \binom{M}{h}\binom{N}{l} \exp\left(-T_{hl}^2\right)$$
$$\times\, \mathbb{E}_0^{hl}\left(\exp\left(\frac{T_{hl}}{a(r_\epsilon)\sqrt{hl}} \sum_k \left(\log\cosh\left(u_k \cdot \frac{x_{11,k}}{\epsilon\sigma_k}\right) - \frac{u_k^2}{2} + \frac{u_k^4}{4}\right)\right)\right).$$

Using Equation (5.11) and applying Lemma A.1 for $\lambda = T_{hl}/(a(r_\epsilon)\sqrt{hl})$ which is $O(1)$, one obtains

$$\mathbb{E}_0\left(\exp\left(\frac{T_{hl}}{a(r_\epsilon)\sqrt{hl}} \sum_k \left(\log\cosh\left(u_k \cdot \frac{x_{11,k}}{\epsilon\sigma_k}\right) - \frac{u_k^2}{2} + \frac{u_k^4}{4}\right)\right)\right) = \exp\left(\frac{T_{hl}^2}{2hl}\left(1 + O\left(a(r_\epsilon) \sup_k w_k^\star\right)\right)\right).$$

Recall that $T_{hl}^2 = 2\log\big(\binom{M}{h}\binom{N}{l}\big)(1+A_\epsilon) + 2\log(mn)$ where $A_\epsilon = o(1)$ by construction. Therefore

$$\begin{aligned}
\mathbb{P}_0\big(\Gamma^C\big) &\le \sum_{\substack{\delta_1 m\le h\le m,\\ \delta_1 n\le l\le n}} \binom{M}{h}\binom{N}{l}\exp\big(-T_{hl}^2\big)\left(\exp\left(\frac{T_{hl}^2}{2hl}\Big(1+O\Big(a(r_\epsilon)\sup_k w_k^\star\Big)\Big)\right)\right)^{hl} \\
&= \sum_{\substack{\delta_1 m\le h\le m,\\ \delta_1 n\le l\le n}} \exp\left(-\frac{T_{hl}^2}{2} + \frac{T_{hl}^2}{2} O\Big(a(r_\epsilon)\sup_k w_k^\star\Big) + \log\left(\binom{M}{h}\binom{N}{l}\right)\right) \\
&\le \exp\left(-\log\left(\binom{M}{h}\binom{N}{l}\right)\cdot A_\epsilon\cdot\left(1-O\Big(a(r_\epsilon)\sup_k w_k^\star\Big)\Big(1+\frac{1}{A_\epsilon}\Big)\right)\right. \\
&\qquad \left. - \log(mn)\Big(1-O\Big(a(r_\epsilon)\sup_k w_k^\star\Big)\Big)\right) = o(1),
\end{aligned}$$

for large enough m, n, M and N, as we have both $A_\epsilon \cdot \log\big(\binom{M}{h}\binom{N}{l}\big) \to \infty$ and $A_\epsilon \cdot /(a(r_\epsilon)\sup_k w_k^\star) \to \infty$ (see (5.12)) and $a(r_\epsilon)\sup_k w_k^\star = o(A_\epsilon) = o(1)$.

2. We have

$$\mathbb{E}_0\big(\tilde{L}_{\overline{\pi}}(\overline{X})\big) = \mathbb{E}_0\left(\frac{1}{\binom{M}{m}\binom{N}{n}} \sum_{\xi\in T_{M,N}(m,n)} \frac{d\mathbb{P}_\xi}{d\mathbb{P}_0}(\overline{X})\mathbb{1}(\Gamma_\xi)\right) = \mathbb{P}_\xi(\Gamma_\xi),$$

which tends to 1 if and only if $\mathbb{P}_\xi(\Gamma_\xi^C) \to 0$. As we can write

$$\mathbb{P}_\xi\big(\Gamma_\xi^C\big) \le \sum_{\substack{\delta_1 m\le h\le m,\\ \delta_1 n\le l\le n}} \sum_{V\subset\xi} \mathbb{P}_V(Y_V > T_{hl}),$$

where $\mathbb{P}_V$ is such that

$$\frac{d\mathbb{P}_V}{d\mathbb{P}_0}(\overline{X}) = \prod_{(i,j)\in A_V\times B_V} \prod_k \exp\left(-\frac{u_k^2}{2}\right)\cosh\left(u_k\frac{x_{ij,k}}{\epsilon\sigma_k}\right) = \exp\left(Y_V a(r_\epsilon)\sqrt{hl} - lh\frac{a^2(r_\epsilon)}{2}\right).$$

Then, applying Lemma A.1, one obtains for any positive λ such that $\lambda = O(1)$,

$$\begin{aligned}
\mathbb{P}_V(Y_V > T_{hl}) &= \mathbb{P}_V\big(Y_V a(r_\epsilon)\sqrt{hl} > T_{hl}a(r_\epsilon)\sqrt{hl}\big) \\
&\le \mathbb{E}_V\big[\exp\big(\lambda Y_V a(r_\epsilon)\sqrt{hl}\big)\big]\exp\big(-\lambda T_{hl}a(r_\epsilon)\sqrt{hl}\big) \\
&= \mathbb{E}_0\big[\exp\big((\lambda+1)Y_V a(r_\epsilon)\sqrt{hl}\big)\big]\exp\left(-lh\frac{a^2(r_\epsilon)}{2} - \lambda T_{hl}a(r_\epsilon)\sqrt{hl}\right) \\
&= \exp\left((\lambda+1)^2 lh\frac{a^2(r_\epsilon)}{2}\big(1+o(1)\big) - lh\frac{a^2(r_\epsilon)}{2} - \lambda T_{hl}a(r_\epsilon)\sqrt{hl}\right).
\end{aligned} \tag{A.2}$$

The minimum value for the right-hand side of (A.2) is

$$\exp\left(-\frac{(T_{hl}-a(r_\epsilon)\sqrt{hl})^2}{2}\big(1+o(1)\big)\right)$$

which is achieved for $\lambda = \frac{T_{hl}}{a(r_\epsilon)\sqrt{hl}} - 1$. Due to $T_{hl}^2 = 2\log\big(\binom{M}{h}\binom{N}{l}\big)(1+A_\epsilon) + 2\log(mn)$ and (A.1), note that λ satisfies $\lambda = O(1)$ and that asymptotically $\frac{T_{hl}^2}{a^2(r_\epsilon)hl} > \frac{2}{2-\delta} > 1$.

In conclusion,

$$\begin{aligned}\mathbb{P}_\xi\big(\Gamma_\xi^C\big) &\le \sum_{\substack{\delta_1 m\le h\le m,\\ \delta_1 n\le l\le n}} \binom{m}{h}\binom{n}{l}\exp\biggl(-\frac{1}{2}\big(T_{hl}-a(r_\epsilon)\sqrt{hl}\big)^2\big(1+o(1)\big)\biggr)\\ &\le \sum_{\substack{\delta_1 m\le h\le m,\\ \delta_1 n\le l\le n}} \binom{m}{h}\binom{n}{l}\exp\biggl(-\frac{c(\delta)}{2}T_{hl}^2\big(1+o(1)\big)\biggr)\end{aligned}$$

for some $c(\delta)>0$ small with δ and where $\binom{m}{h}\binom{n}{l}\exp(-\frac{c(\delta)}{2}T_{hl}^2(1+o(1)))$ is bounded by

$$\exp\biggl(h\log\biggl(\frac{me}{h}\biggr)+l\log\biggl(\frac{ne}{l}\biggr)-\frac{c(\delta)}{2}\biggl(2h\log\biggl(\frac{M}{h}\biggr)+2l\log\biggl(\frac{N}{l}\biggr)\biggr)(1+A_\epsilon)\big(1+o(1)\big)-\log(mn)\biggr).$$

We see that

$$h\log\biggl(\frac{me}{h}\biggr)\le m\log\biggl(\frac{e}{\delta_1}\biggr),\quad \forall h:\delta_1 m\le h\le m$$

and

$$h\log\biggl(\frac{M}{h}\biggr)(1+A_\epsilon)\ge c_1\delta_1 m\log\biggl(\frac{M}{m}\biggr)\text{ asymptotically,}\quad \text{for some } c_1>0.$$

The same occurs for terms in l, N and n. Therefore, it implies that asymptotically for some $\tilde{c}_1(\delta)>0$ and $\tilde{c}_2(\delta)>0$

$$\mathbb{P}_\xi\big(\Gamma_\xi^C\big)\le mn\exp\biggl(-\tilde{c}_1(\delta)m\log\biggl(\frac{M}{m}\biggr)-\tilde{c}_2(\delta)n\log\biggl(\frac{N}{n}\biggr)-\frac{c(\delta)}{2}\log(nm)\biggr)=o(1).$$

3. We have, for $\xi_1=\mathbb{1}((i,j)\in A_1\times B_1)$ and $\xi_2=\mathbb{1}((i,j)\in A_2\times B_2)$,

$$\mathbb{E}_0\big(\tilde{L}_{\overline{\pi}}^2(\overline{X})\big)=\frac{1}{(\binom{M}{m}\binom{N}{n})^2}\sum_{\xi_1}\sum_{\xi_2} g\big(h(\xi_1,\xi_2),l(\xi_1,\xi_2)\big),$$

where

$$\begin{aligned}&g\big(h(\xi_1,\xi_2),l(\xi_1,\xi_2)\big)\\ &\quad=\prod_{(i_1,j_1)\in A_1\times B_1}\prod_{(i_2,j_2)\in A_2\times B_2}\prod_k\exp\biggl(-\frac{\theta_k^{\star 2}}{\epsilon^2\sigma_k^2}\biggr)\\ &\qquad\cdot\mathbb{E}_0\biggl(\cosh\biggl(x_{i_1j_1,k}\frac{\theta_k^\star}{\epsilon^2\sigma_k^2}\biggr)\cosh\biggl(x_{i_2j_2,k}\frac{\theta_k^\star}{\epsilon^2\sigma_k^2}\biggr)\mathbb{1}(\Gamma_{\xi_1}\cap\Gamma_{\xi_2})\biggr)\\ &\quad=\prod_{(i,j)\in(A_1\cap A_2)\times(B_1\cap B_2)}\prod_k\exp\biggl(-\frac{\theta_k^{\star 2}}{\epsilon^2\sigma_k^2}\biggr)\mathbb{E}_0\biggl(\cosh^2\biggl(x_{ij,k}\frac{\theta_k^\star}{\epsilon^2\sigma_k^2}\biggr)\mathbb{1}(\Gamma_{\xi_1}\cap\Gamma_{\xi_2})\biggr)\end{aligned}$$

and the function g depends on the sets A_1, A_2 and B_1, B_2 only through the number $h(\xi_1,\xi_2)$ of common rows of A_1 and A_2 and the number $l(\xi_1,\xi_2)$ of common columns of B_1 and B_2. For the sake of simplicity, denote $h:=h(\xi_1,\xi_2)$ and $l:=l(\xi_1,\xi_2)$. After some combinatorics we can write

$$\mathbb{E}_0\big(\tilde{L}_{\overline{\pi}}^2(\overline{X})\big)=\mathbb{E}\big(g(H,L)\big),$$

where conditionally on ξ_1, H and L are independent random variables having hypergeometric distribution $\mathcal{HG}(M,m,m)$ and $\mathcal{HG}(N,n,n)$, respectively. Let us see that, for any $0 \le h \le m$ and $0 \le l \le n$,

$$\begin{aligned}
\log\bigl(g(h,l)\bigr) &\le \sum_{(i,j)\in(A_1\cap A_2)\times(B_1\cap B_2)} \log\biggl(\prod_k \exp\biggl(-\frac{\theta_k^{\star 2}}{\epsilon^2\sigma_k^2}\biggr)\mathbb{E}_0\biggl(\cosh^2\biggl(x_{ij,k}\frac{\theta_k^\star}{\epsilon^2\sigma_k^2}\biggr)\biggr)\biggr) \\
&= hl\log\biggl(\prod_k \exp\biggl(-\frac{\theta_k^{\star 2}}{\epsilon^2\sigma_k^2}\biggr)\frac{1}{2}\biggl(\exp\biggl(\frac{2\theta_k^{\star 2}}{\epsilon^2\sigma_k^2}\biggr)+1\biggr)\biggr) \\
&= hl\log\biggl(\prod_k \cosh\biggl(\frac{\theta_k^{\star 2}}{\epsilon^2\sigma_k^2}\biggr)\biggr) = hl\cdot D.
\end{aligned}$$

Therefore, $\mathbb{E}(g(H,L)) \le \mathbb{E}(e^{HL\cdot D})$ for $D = a^2(r_\epsilon)(1+o(1))$ (see equation (5.8)).

We shall split $\mathbb{E}(g(H,L))$ into the sum $I_1 + I_2$, where

$$\begin{aligned}
I_1 &= \mathbb{E}\bigl(g(H,L)\cdot \mathbb{1}(HD<1)\bigr), \\
I_2 &= \mathbb{E}\bigl(g(H,L)\cdot \mathbb{1}(HD\ge 1)\bigr).
\end{aligned}$$

3.1. For I_1, we act exactly as for (5.9); again due to Proposition 20.6 in [1] and given $\mathcal{X}_1$, there exist two Binomial random variables $Y_1 \sim B(m,p)$ and $Y_2 \sim B(n,q)$ such that $\mathbb{E}^{\mathcal{X}_1}_{\mathcal{X}_1,\mathcal{X}_2}(Y_1|H) = H$ and $\mathbb{E}^{\mathcal{X}_1}_{\mathcal{X}_1,\mathcal{X}_2}(Y_2|L) = L$; then applying twice Jensen Inequality and using the generating function of Binomial random variables, we get

$$\begin{aligned}
I_1 &\le \mathbb{E}\bigl(\mathbb{E}\bigl(e^{LHD}\mathbb{1}(HD<1)|H\bigr)\bigr) \\
&= \mathbb{E}\bigl(\mathbb{E}^{\mathcal{X}_1}\mathbb{E}\bigl(e^{\mathbb{E}^{\mathcal{X}_1}_{\mathcal{X}_1,\mathcal{X}_2}(Y_2|L)HD}|H\bigr)\mathbb{1}(HD<1)\bigr) \\
&\le \mathbb{E}\bigl(\mathbb{E}\bigl(e^{Y_2HD}|H\bigr)\mathbb{1}(HD<1)\bigr) \\
&\le \mathbb{E}\bigl(\bigl(1+q\bigl(e^{HD}-1\bigr)\bigr)^n\mathbb{1}(HD<1)\bigr) \\
&\le \mathbb{E}\bigl(\exp\bigl(CnqHD\bigl(1+o(1)\bigr)\bigr)\bigr) \\
&= \mathbb{E}\bigl(\exp\bigl(CnqE(Y_1|H)D\bigl(1+o(1)\bigr)\bigr)\bigr) \\
&\le E\bigl(\exp\bigl(CnqY_1D\bigl(1+o(1)\bigr)\bigr)\bigr) = \bigl(1+p\bigl(e^{CnqD(1+o(1))}-1\bigr)\bigr)^m,
\end{aligned}$$

for some constant $C>0$. By assumption (3.2) and relations (A.1) and (5.8), $Dn \asymp \log(p^{-1})$, which implies that $Dnq \asymp (q\log(p^{-1}))$ and this is an $o(1)$ by assumption (3.1). Then, by assumption (3.4), $I_1 \le \exp(CmnpqD(1+o(1))) = 1+o(1)$.

3.2. The rest of the section is devoted to the proof of $I_2 = o(1)$. We shall further split the expected value into the sum of $I_{21}+I_{22}$, where

$$\begin{aligned}
I_{21} &= \mathbb{E}\bigl(g(H,L)\cdot \mathbb{1}(HD\ge 1)\cdot \mathbb{1}(L<n\delta_1)\bigr), \\
I_{22} &= \mathbb{E}\bigl(g(H,L)\cdot \mathbb{1}(HD\ge 1)\cdot \mathbb{1}(L\ge n\delta_1)\bigr),
\end{aligned}$$

for some fixed $\delta_1>0$, small enough such that $Dn\delta_1 \le \log(p^{-1})/2$ and that $Dm\delta_1 \le \log(q^{-1})/2$.

3.2.1. On the one hand

$$\begin{aligned}
I_{21} &\le \sum_{D^{-1}<h\le m, 0\le l<n\delta_1} e^{hlD}\mathbb{P}_{\mathcal{HG}(M,m,m)}(H=h)\mathbb{P}_{\mathcal{HG}(N,n,n)}(L=l) \\
&\le \sum_{D^{-1}<h\le m, 0\le l<n\delta_1} e^{h(lD-\log(p^{-1})(1+o(1)))},
\end{aligned}$$

as $\mathbb{P}_{\mathcal{HG}(N,n,n)}(L=l) \le 1$ and by using Lemma 5.3 in Butucea and Ingster [4] for $\log(\mathbb{P}_{\mathcal{HG}(M,m,m)}(H=h)) \le h\log(p)(1+o(1))$. Note that in our case, we can apply Lemma 5.3 in Butucea and Ingster [4] due to assumption (3.1). Now, under the constraints in the sum, $lD \le Dn\delta_1 \le \log(p^{-1})/2$. This implies that $h(lD - \log(p^{-1})(1+o(1))) \le -h\log(p^{-1})(1/2+o(1)) \le -D^{-1}\log(p^{-1})(1/2+o(1)) \asymp -n$. Therefore,

$$I_{21} \le mne^{-Bn}, \quad \text{for some fixed } B > 0,$$

and this is an $o(1)$.

3.2.2. We can also split I_{22} into the sum of $I_{221} + I_{222}$, where

$$I_{221} = \mathbb{E}\big(g(H,L)\cdot \mathbb{1}(HD \ge 1, H < m\delta_1)\cdot \mathbb{1}(L \ge n\delta_1)\big),$$
$$I_{222} = \mathbb{E}\big(g(H,L)\cdot \mathbb{1}(HD \ge 1, H \ge m\delta_1)\cdot \mathbb{1}(L \ge n\delta_1)\big).$$

It is easy to check that $I_{221} = o(1)$ as we previously did for I_{21}, except that Lemma 5.3 in Butucea and Ingster [4] is used to control the term $\mathbb{P}_{\mathcal{HG}(N,n,n)}(L=l)$.

On the other hand, we can write

$$I_{222} = \mathbb{E}\big(g(H,L)\mathbb{1}(\mathcal{H})\big), \quad \text{where } \mathcal{H} = \big\{(h,l), m\delta_1 \le h \le m, n\delta_1 \le l \le n\big\}.$$

I_{222} is the only term for which the truncation is really required to prove that $I_{222} = o(1)$. Note that under the event $\mathcal{H}$ we have $T_{hl}^2 = (2+\delta)(h/m \cdot m\log(p^{-1}) + l/n \cdot n\log(q^{-1})) \ge \delta_1 T_{mn}^2$.

We divide again the set $\mathcal{H}$ in disjoint sets

$$\mathcal{H}_1 = \left\{(h,l) \in \mathcal{H} : T_{hl}^2 > 2T_{mn}^2 \frac{hl}{mn}\right\} \quad \text{and} \quad \mathcal{H}_2 = \left\{(h,l) \in \mathcal{H} : T_{hl}^2 \le 2T_{mn}^2 \frac{hl}{mn}\right\}.$$

Let us go back to $L_{\overline{\pi}}(\overline{X})$ and rewrite it as follows

$$= \frac{1}{\binom{M}{m}\binom{N}{n}} \sum_{\xi \in T_{M,N}(m,n)} \exp\left(\sum_{(i,j)\in A_\xi \times B_\xi} \sum_k \left(\log\cosh\left(\frac{x_{ij,k}}{\epsilon\sigma_k}u_k\right) - \frac{u_k^2}{2}\right)\right)$$
$$= \frac{\exp(-a^2(r_\epsilon)mn/2)}{\binom{M}{m}\binom{N}{n}} \sum_{\xi \in T_{M,N}(m,n)} \exp\left(\sum_{(i,j)\in A_\xi \times B_\xi} \sum_k \left(\log\cosh\left(\frac{x_{ij,k}}{\epsilon\sigma_k}u_k\right) - \frac{u_k^2}{2} + \frac{u_k^4}{4}\right)\right).$$

Now, we give a tighter upper bound for $g(h,l)$ than the one used for I_1. Using the same notation as to define Y_V in (5.10) and for any matrix ξ, we define the random variable $Y_\xi = \frac{1}{a(r_\epsilon)\sqrt{hl}} \sum_{(i,j)\in A_\xi \times B_\xi} \sum_k (\log\cosh(\frac{x_{ij,k}}{\epsilon\sigma_k}u_k) - \frac{u_k^2}{2} + \frac{u_k^4}{4})$. Then, we write

$$\begin{aligned} g(h,l) &= e^{-a^2(r_\epsilon)mn}\mathbb{E}_0\big(e^{a(r_\epsilon)\sqrt{mn}(Y_{\xi_1}+Y_{\xi_2})}\mathbb{1}(\varGamma_{\xi_1} \cap \varGamma_{\xi_2})\big) \\ &\le e^{-a^2(r_\epsilon)mn}\mathbb{E}_0\big(e^{a(r_\epsilon)\sqrt{mn}(Y_{\xi_1}+Y_{\xi_2})}\mathbb{1}(Y_{\xi_1} \le T_{mn}, Y_{\xi_2} \le T_{mn})\big) \\ &\le e^{-a^2(r_\epsilon)mn + 2T_{mn}J}\mathbb{E}_0\big(e^{(a(r_\epsilon)\sqrt{mn}-J)(Y_{\xi_1}+Y_{\xi_2})}\big), \end{aligned} \tag{A.3}$$

for some $J > 0$ that we will choose later on. In order to deal with I_{222}, we keep in mind that we consider only submatrices ξ_1 and ξ_2 having h common rows and l common columns, such that $(h,l) \in \mathcal{H}$. Denote by V the submatrix of common rows and columns for ξ_1 and ξ_2, that is

$$V = \mathbb{1}\big((i,j) \in (A_{\xi_1} \times B_{\xi_1}) \cap (A_{\xi_2} \times B_{\xi_2})\big),$$

and by $V_1 = \xi_1 - V$ (respectively $V_2 = \xi_2 - V$). Therefore,

$$\sqrt{mn}(Y_{\xi_1} + Y_{\xi_2}) = \sqrt{mn - hl}(Y_{V_1} + Y_{V_2}) + 2\sqrt{hl}Y_V.$$

Replace this into the equation (A.3) and get by Lemma A.1

$$\begin{aligned}
I_{222} &\le \sum_{(h,l)\in\mathcal{H}} \exp\biggl(-a^2(r_\epsilon)mn + 2T_{mn}J + \bigl(a(r_\epsilon)\sqrt{mn} - J\bigr)^2\biggl(1+\frac{hl}{mn}\biggr)\biggr)\mathbb{P}_{\mathcal{HG}(M,m,m)}(h)\mathbb{P}_{\mathcal{HG}(N,n,n)}(l) \\
&\le \sum_{(h,l)\in\mathcal{H}} \exp\biggl(-a^2(r_\epsilon)mn + 2T_{mn}a(r_\epsilon)\sqrt{mn} - \frac{T_{mn}^2}{1+hl/(mn)} - \bigl(h\log\bigl(p^{-1}\bigr) + l\log\bigl(q^{-1}\bigr)\bigr)\bigl(1+o(1)\bigr)\biggr) \\
&\le \sum_{(h,l)\in\mathcal{H}} \exp\biggl(-\bigl(a(r_\epsilon)\sqrt{mn} - T_{mn}\bigr)^2 + \frac{T_{mn}^2 hl}{mn+hl} - \frac{T_{hl}^2}{2}\bigl(1+o(1)\bigr)\biggr),
\end{aligned}$$

for $a(r_\epsilon)\sqrt{mn} - J = T_{mn}/(1+hl/(mn))$. Note that, for $\delta > 0$ small enough, there exists $\delta_2 > 0$ such that $(a(r_\epsilon)\sqrt{mn} - T_{mn})^2 \ge \delta_2 T_{mn}^2$. Moreover, for $(h,l) \in \mathcal{H}_1$,

$$\begin{aligned}
\frac{T_{mn}^2 hl}{mn+hl} - \bigl(h\log\bigl(p^{-1}\bigr) + l\log\bigl(q^{-1}\bigr)\bigr)\bigl(1+o(1)\bigr) &\le \frac{T_{mn}^2 hl}{mn+hl} - \frac{T_{hl}^2(1+o(1))}{2} \\
&\le T_{mn}^2 \frac{hl}{mn}\biggl(\frac{1}{1+hl/(mn)} - 1\biggr) + o\bigl(T_{mn}^2\bigr),
\end{aligned}$$

which is negative and asymptotically $O(T_{mn}^2)$. This implies that $I_{222} = o(1)$ over the set $\mathcal{H}_1$.

Finally, we give a yet slightly different upper bound for $g(h,l)$ in order to deal with I_{222} when (h,l) belongs to $\mathcal{H}_2$. Again with the submatrix $V = \mathbb{1}((i,j) \in (A_{\xi_1} \times B_{\xi_1}) \cap (A_{\xi_2} \times B_{\xi_2}))$, one has,

$$\begin{aligned}
g(h,l) &\le e^{-a^2(r_\epsilon)mn}\mathbb{E}_0\bigl(e^{a(r_\epsilon)\sqrt{mn}(Y_{\xi_1}+Y_{\xi_2})}\mathbb{1}(Y_V \le T_{hl})\bigr) \\
&\le e^{-a^2(r_\epsilon)hl}\mathbb{E}_0\bigl(e^{2a(r_\epsilon)\sqrt{hl}Y_V}\mathbb{1}(Y_V \le T_{hl})\bigr) \\
&\le e^{-a^2(r_\epsilon)hl + T_{hl}J}\mathbb{E}_0\bigl(e^{(2a(r_\epsilon)\sqrt{hl}-J)Y_V + J(Y_V - T_{hl})}\mathbb{1}(Y_V \le T_{hl})\bigr) \\
&\le e^{-a^2(r_\epsilon)hl + T_{hl}J + (2a(r_\epsilon)\sqrt{hl}-J)^2/2}.
\end{aligned}$$

Take $J = 2a(r_\epsilon)\sqrt{hl} - T_{hl}$ which is indeed positive for (h,l) in $\mathcal{H}_2$ and obtain

$$g(h,l) \le \exp\biggl(-a^2(r_\epsilon)hl + 2a(r_\epsilon)\sqrt{hl}T_{hl} - \frac{T_{hl}^2}{2}\biggr).$$

Moreover, denote $D_{hl}^2 = h\log(p^{-1}) + l\log(q^{-1})$ and see that $D_{mn}^2 hl/(mn) \le D_{hl}^2$. We get

$$\begin{aligned}
I_{222} &\le \sum_{(h,l)\in\mathcal{H}} \exp\biggl(-a^2(r_\epsilon)hl + 2a(r_\epsilon)\sqrt{hl}T_{hl} - \frac{T_{hl}^2}{2}\biggr)\mathbb{P}_{\mathcal{HG}(M,m,m)}(h)\mathbb{P}_{\mathcal{HG}(N,n,n)}(l) \\
&\le \sum_{(h,l)\in\mathcal{H}} \exp\biggl(-\bigl(a(r_\epsilon)\sqrt{h}l - T_{hl}\bigr)^2 + \frac{T_{hl}^2}{2} - D_{hl}^2\bigl(1+o(1)\bigr)\biggr) \\
&\le \sum_{(h,l)\in\mathcal{H}} \exp\biggl(-\frac{\delta^2}{8}D_{hl}^2 + o(1)D_{hl}^2\biggr) = o(1),
\end{aligned}$$

where the last inequality derived from the following relations

$$\begin{cases}
T_{hl} - a(r_\epsilon)\sqrt{hl} \ge \sqrt{2}D_{hl}(1-\sqrt{1-\delta/2}) + o(D_{hl}) \ge \frac{\sqrt{2}}{4}D_{hl} + o(D_{hl}), \\
\frac{T_{hl}^2}{2} - D_{hl}^2(1+o(1)) = D_{hl}^2(A_\epsilon + o(1)).
\end{cases}$$

A.2. *Proof of Lemma* 5.1

For the sake of simplicity, we omit in this part the indices i and j so that $t_{ij,w^\star}$ and $\eta_{ij,k}$ are denoted by $t_{w^\star}$ and η_k, respectively.

Under H_0, observe that $t_{w^\star} = \sum_k w_k^\star(\eta_k^2 - 1)$, with $\eta_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$. Using the fact that $\mathbb{E}(e^{t\eta_k^2}) = \frac{1}{(1-2t)^{1/2}}$, for $t \le \frac{1}{2}$, we obtain for λ such that $\lambda \sup_k w_k^\star = o(1)$,

$$\begin{aligned}
\mathbb{E}_0\bigl(\exp(\lambda t_{w^\star})\bigr) &= \prod_{k\in\mathbb{Z}} \exp\biggl(-\lambda w_k^\star - \frac{1}{2}\log\bigl(1 - 2\lambda w_k^\star\bigr)\biggr) \\
&= \exp\biggl(\lambda^2 \sum_k \bigl(w_k^\star\bigr)^2\bigl(1 + o(1)\bigr)\biggr) \\
&= \exp\biggl(\frac{\lambda^2}{2}\bigl(1 + o(1)\bigr)\biggr),
\end{aligned}$$

where the last equality holds since $\sum_k (w_k^\star)^2 = \frac{1}{2}$.

A.3. *Proof of Proposition* 2.1

These computations can be found in Ingster and Suslina [17], but we give the sketch of proof for the convenience of the reader.

Let us change variables in problem (2.2), by defining $v_k = \frac{\theta_k^2}{\sigma_k^2\sqrt{2}}$, for all $k \in \mathbb{Z}$. We have $\{\theta_k\}_k$ belongs to $\Sigma(\tau, r_\epsilon)$ if and only if $\{v_k\}_k$ belongs to $\tilde{\Sigma}(\tau, r_\epsilon)$, where

$$\tilde{\Sigma}(\tau, r_\epsilon) = \biggl\{\{v_k\}_k \in l_1(\mathbb{Z}) : v_k \ge 0;\ (2\pi)^{2\tau} \sum_{k\in\mathbb{Z}} |k|^{2\tau}\sigma_k^2 v_k \le \frac{1}{\sqrt{2}};\ \sum_{k\in\mathbb{Z}} v_k\sigma_k^2 \ge \frac{r_\epsilon^2}{\sqrt{2}}\biggr\}.$$

The problem (2.2) is equivalent to

$$\begin{aligned}
&\frac{\sqrt{2}}{\epsilon^2} \sup_{\{w_k\}_k:\sum_k w_k^2 = 1/2, w_k \ge 0}\ \inf_{\{v_k\}_k \in \tilde{\Sigma}(\tau, r_\epsilon)} \sum_k w_k v_k \\
&\quad = \frac{\sqrt{2}}{\epsilon^2} \sup_{\{w_k\}_k:\sum_k w_k^2 \le 1/2, w_k \ge 0}\ \inf_{\{v_k\}_k \in \tilde{\Sigma}(\tau, r_\epsilon)} \sum_k w_k v_k \\
&\quad = \frac{\sqrt{2}}{\epsilon^2} \inf_{\{v_k\}_k \in \tilde{\Sigma}(\tau, r_\epsilon)}\ \sup_{\{w_k\}_k:\sum_k w_k^2 \le 1/2, w_k \ge 0} \sum_k w_k v_k,
\end{aligned}$$

by the minimax theorem on convex sets. Now, use the Cauchy–Schwarz inequality to see that

$$\sqrt{2} \sup_{\{w_k\}_k:\sum_k w_k^2 \le 1/2, w_k \ge 0} \sum_k w_k v_k = \biggl(\sum_k v_k^2\biggr)^{1/2}$$

and the equality holds for $w_k = v_k(2\sum_k v_k^2)^{-1/2}$. Since we denoted by $\epsilon^2 a(r_\epsilon) = (\sum_k (v_k^\star)^2)^{1/2}$ we get $w_k^\star = v_k^\star/(\sqrt{2}\epsilon^2 a(r_\epsilon))$, which is equivalent to

$$w_k^\star = \frac{(\theta_k^\star)^2}{2\sigma_k^2\epsilon^2 a(r_\epsilon)}, \quad \text{for all } k \in \mathbb{Z}.$$

It follows that solving the problem (2.2) reduces to solve the optimization program

$$\inf_{\{v_k\}_k, v_k \geq 0} \sum_k v_k^2 + \lambda_1 \left(\sum_k (2\pi|k|)^{2\tau} \sigma_k^2 v_k - \frac{1}{\sqrt{2}} \right) - \lambda_2 \left(\sum_k \sigma_k^2 v_k - \frac{r_\epsilon^2}{\sqrt{2}} \right).$$

By the Lagrangian multipliers rules, one gets for $\lambda_1 \in \mathbb{R}$ and $\lambda_2 \in \mathbb{R}$ the following system of equations

$$\begin{cases} 2v_k + \lambda_1 \sqrt{2}(2\pi)^{2\tau}(|k|^{2\tau}\sigma_k^2) - \lambda_2 \sqrt{2}\sigma_k^2 = 0, & \text{for all } k \in \mathbb{Z}, \\ \sqrt{2}(2\pi)^{2\tau} \sum_k v_k |k|^{2\tau} \sigma_k^2 = 1, \\ \sqrt{2} \sum_k v_k \sigma_k^2 = r_\epsilon^2. \end{cases}$$

Put, for all $k \in \mathbb{Z}$, $v_k = v\sigma_k^2(1 - (\frac{|k|}{T})^{2\tau})_+$, where $v = \frac{\lambda_2}{\sqrt{2}}$, $T = \frac{1}{2\pi}(\frac{\lambda_2}{\lambda_1})^{1/(2\tau)}$ and $(x)_+ = \max(0, x)$.

We evaluate the solution of the previous system as T goes to infinity. Using $\sigma_k \sim |k|^s$ for $|k|$ large enough and some $s > 0$, the last two equations in the previous system become

$$\begin{cases} \kappa_2 v T^{2\tau+4s+1} \sim 1, \\ \kappa_1 v T^{4s+1} \sim r_\epsilon^2, \end{cases}$$

that gives

$$T \sim \left(\frac{\kappa_1}{\kappa_2}\right)^{1/(2\tau)} r_\epsilon^{-\frac{1}{\tau}} \quad \text{and} \quad v \sim \frac{1}{\kappa_1}\left(\frac{\kappa_2}{\kappa_1}\right)^{(4s+1)/(2\tau)} r_\epsilon^{2+(4s+1)/\tau}.$$

Note that $T \to \infty$ provided that $r_\epsilon \to 0$. It further gives

$$a^2(r_\epsilon)\epsilon^4 = \sum_{|k| \leq T} (v_k^\star)^2 \sim 2v^2 T^{4s+1} \kappa_3 \sim c(\tau, s)^2 r_\epsilon^{4+(4s+1)/\tau}.$$

Finally, it is straightforward that

$$\begin{aligned} \sup_k w_k^\star &\leq \frac{v}{a(r_\epsilon)\epsilon^2} \max_{0 \leq |k| \leq T} \sigma_k^2 \left(1 - \left(\frac{|k|}{T}\right)^{2\tau}\right) \\ &\leq \frac{v}{a(r_\epsilon)\epsilon^2} \sigma_T^2 \asymp r_\epsilon^{(4s+2\tau+1)/\tau - (4s+4\tau+1)/(2\tau) - (2s)/\tau} = r_\epsilon^{1/(2\tau)} \xrightarrow{r_\epsilon \to 0} 0. \end{aligned}$$

References

[1] D. Aldous. Exchangeability and related topics. In *École d'Été de Probabilités Saint-Flour XIII* 1–198. *Lecture Notes in Mathematics* **1117**. Springer, Berlin, 1985. MR0883646

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

[3] C. Butucea. Goodness-of-fit testing and quadratic functional estimation from indirect observations. *Ann. Statist.* **35** (2007) 1907–1930. MR2363957

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

[5] L. Cavalier, G. K. Golubev, D. Picard and A. B. Tsybakov. Oracle inequalities for inverse problems. Dedicated to the memory of Lucien Le Cam. *Ann. Statist.* **30** (2002) 843–874. MR1922543

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

[7] G. Gayraud and Y. I. Ingster. Detection of sparse additive variable functions. *Electron. J. Stat.* **6** (2012) 1409–1448. MR2988453

[8] Y. I. Ingster. Asymptotically minimax hypothesis testing for nonparametric alternatives. I. *Math. Methods Statist.* **2** (1993) 85–114. MR1257978

[9] Yu. I. Ingster. Asymptotically minimax hypothesis testing for nonparametric alternatives. II. *Math. Methods of Statist.* **2** (1993) 171–189. MR1257983

[10] Yu. I. Ingster. Asymptotically minimax hypothesis testing for nonparametric alternatives. III. *Math. Methods of Statist.* **2** (1993) 249–268. MR1259685

[11] Y. I. Ingster. Some problems of hypothesis testing leading to infinitely divisible distributions. *Math. Methods Statist.* **6** (1997) 47–69. MR1456646

[12] Y. I. Ingster, T. Sapatinas and I. A. Suslina. Minimax nonparametric testing in a problem related to the Radon transform. *Math. Methods Statist.* **20** (2011) 347–364. MR2886641

[13] Y. I. Ingster, T. Sapatinas and I. A. Suslina. Minimax signal detection in ill-posed inverse problems. *Ann. Statist.* **40** (2012) 1524–1549. MR3015034

[14] Y. I. Ingster and N. Stepanova. Estimation and detection of functions from weighted tensor product spaces. *Math. Methods Statist.* **18** (2009) 310–340. MR2608165

[15] Y. I. Ingster and N. Stepanova. Estimation and detection of functions from anisotropic Sobolev classes. *Electron. J. Stat.* **5** (2011) 484–506. MR2813552

[16] Y. I. Ingster and I. A. Suslina. On a detection of a signal of known shape in multichannel system. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)* **294** (2002) 88–112. Transl. *J. Math. Sci.* **127** (2005) 1723–1736. MR1976749

[17] Y. I. Ingster and I. A. Suslina. *Nonparametric Goodness-of-Fit Testing Under Gaussian Models*. *Lectures Notes in Statistics* **169**. Springer, New York, 2003. MR1991446

[18] B. Laurent, J.-M. Loubes and C. Marteau. Signal detection for inverse problems in a multidimensional framework. *Math. Methods Statist.* **23** (2014) 279–305. MR3295060

[19] 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

[20] P. Ravikumar, J. Lafferty, H. Liu and L. Wasserman. Sparse additive models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** (2009) 1009–1030. MR2750255

[21] X. Sun and A. B. Nobel. On the maximal size of large-average and ANOVA-fit submatrices in a Gaussian random matrix. *Bernoulli* **20** (2013) 275–294. MR3019495

[22] A. A. Shabalin, V. J. Weigman, C. M. Perou and A. B. Nobel. Finding large average submatrices in high dimensional data. *Ann. Appl. Stat.* **3** (2009) 985–1012. MR2750383

[23] C. Stone. Additive regression and other nonparametric models. *Ann. Statist.* **13** (1985) 689–705. MR0790566

[24] A. B. Tsybakov. *Introduction to Nonparametric Statistics*. *Springer Series in Statistics*. Springer, New York, 2009. MR2724359

[25] G. Walther. Optimal and fast detection of spatial clusters with scan statistics. *Ann. Statist.* **38** 1010–1033. MR2604703