Bernoulli **19**(5B), 2013, 2652–2688
DOI: 10.3150/12-BEJ470

Detection of a sparse submatrix of a high-dimensional noisy matrix

CRISTINA BUTUCEA[1] and YURI I. INGSTER[2]

[1]*Université Paris-Est, LAMA (UMR 8050), UPEMLV, UPEC, CNRS, F-77464, Marne-la-Vallée, France and CREST, Timbre J340 3, av. Pierre Larousse, 92240 Malakoff Cedex, France.*
E-mail: cristina.butucea@univ-mlv.fr
[2]*St. Petersburg Electrotechnical University, 5, Prof. Popov str., 197376 St. Petersburg, Russia.*
E-mail: yurii_ingster@mail.ru

We observe a $N \times M$ matrix $Y_{ij} = s_{ij} + \xi_{ij}$ with $\xi_{ij} \sim \mathcal{N}(0, 1)$ i.i.d. in i, j, and $s_{ij} \in \mathbb{R}$. We test the null hypothesis $s_{ij} = 0$ for all i, j against the alternative that there exists some submatrix of size $n \times m$ with significant elements in the sense that $s_{ij} \ge a > 0$. We propose a test procedure and compute the asymptotical detection boundary a so that the maximal testing risk tends to 0 as $M \to \infty$, $N \to \infty$, $p = n/N \to 0$, $q = m/M \to 0$. We prove that this boundary is asymptotically sharp minimax under some additional constraints. Relations with other testing problems are discussed. We propose a testing procedure which adapts to unknown (n, m) within some given set and compute the adaptive sharp rates. The implementation of our test procedure on synthetic data shows excellent behavior for sparse, not necessarily squared matrices. We extend our sharp minimax results in different directions: first, to Gaussian matrices with unknown variance, next, to matrices of random variables having a distribution from an exponential family (non-Gaussian) and, finally, to a two-sided alternative for matrices with Gaussian elements.

Keywords: detection of sparse signal; minimax adaptive testing; minimax testing; random matrices; sharp detection bounds

1. Introduction

We observe a high-dimensional random matrix and we want to test the occurrence of a particular submatrix of much smaller size, which has elements with expected values larger than some threshold. We assume that the entries of the matrix are independent, identically distributed (i.i.d.) random variables but some underlying phenomenon can increase significantly the expected value of the random variables in the submatrix.

We have the observations that form an $N \times M$ matrix $\mathbf{Y} = \{Y_{ij}\}_{i=1,\dots,N, j=1,\dots,M}$:

$$Y_{ij} = s_{ij} + \sigma \xi_{ij}, \qquad i = 1, \dots, N, j = 1, \dots, M, \tag{1.1}$$

where $\sigma > 0$, $\{\xi_{ij}\}$ are i.i.d. random variables and $s_{ij} \in \mathbb{R}$, for all $i \in \{1, \dots, N\}$, $j \in \{1, \dots, M\}$. In the first part of the paper, the errors ξ_{ij} are assumed to have standard Gaussian law and σ is assumed to be known. Without loss of generality, we take $\sigma = 1$ in this case. At the end of the paper, we extend our results in different directions, as discussed later on. We test the null hypothesis that all elements of the matrix $\mathbf{Y}$ are i.i.d., standard Gaussian random variables

$\mathcal{N}(0,1)$, that is

$$H_0: \quad s_{ij} = 0 \qquad \forall i = 1, \dots, N, \ j = 1, \dots, M. \tag{1.2}$$

The alternative under consideration will correspond to $n \times m$-submatrices of sizes $n \in \{1, \dots, N\}$, $m \in \{1, \dots, M\}$ with large enough entries. Let

$$A \subset \{1, \dots, N\}, \qquad \#(A) = n, \qquad B \subset \{1, \dots, M\}, \qquad \#(B) = m, \qquad C = A \times B, \tag{1.3}$$

and let $\mathcal{C}_{nm}$ be the collection of all subsets C of the form (1.3). The set $\mathcal{C}_{nm}$ corresponds to the collection of all $n \times m$ submatrices in $N \times M$ matrix. For $a > 0$, which may depend on n, m, N and M. We consider the alternative

$$H_1: \quad \exists C \in \mathcal{C}_{nm} \text{ such that } s_{ij} = 0 \text{ if } (i,j) \notin C \text{ and } s_{ij} \geq a \text{ if } (i,j) \in C \tag{1.4}$$

(in the Remark 2.1 below we discuss that a slightly larger alternative can be considered). The components of the matrix $\mathbf{Y}$ are independent under the alternative as well. Denote by P_S the probability measure that corresponds to observations (1.1) with matrix $S = \{s_{ij}\}$ and by E_S the expected value with respect to the measure P_S.

Let $\mathcal{S}_{nm,a}$ be the collection of all matrices $S = S_C$ that satisfy (1.4).

We discuss here only right-hand side alternatives, but, obviously, left-hand side alternatives can be treated the same way for variables $-Y_{ij}$ instead of Y_{ij}.

We extend our results to three different setups and sketch the proofs of the results. First, we consider errors having Gaussian distribution with unknown variance σ^2. We also consider other settings where the Y_{ij}'s come from an exponential family. Finally, in the initial case of Gaussian errors with known variance, we consider a two-sided alternative of our test problem.

We are interested here in sparse matrices, that is, the case when n is much smaller than N and m is much smaller than M.

Sparsity assumptions were introduced for vectors. Estimation as well as hypothesis testing for vectors were thoroughly studied in the literature, see, for example, Bickel, Ritov and Tsybakov [5] and references therein and Donoho and Jin [6].

In the context of matrices, different sparsity assumptions can be imagined. For example, matrix completion for low rank matrices with the nuclear norm penalization has been studied by Koltchinskii, Lounici and Tsybakov [9]. Other results will be discussed later on.

We study the hypothesis testing problem under a minimax setting. A test is any measurable function of the observations, $\psi = \psi(\{Y_{ij}\})$ taking values in $[0,1]$. For such a test $\psi = \psi(\{Y_{ij}\})$, we denote the probability of type-I error, the probability of type-II error under simple alternative and the maximal probability of type-II error over the set $\mathcal{S}_{nm,a}$ by

$$\alpha(\psi) = E_0\psi, \qquad \beta(\psi, S) = E_S(1-\psi), \qquad \beta_{nm,a}(\psi) = \sup_{S \in \mathcal{S}_{nm,a}} \beta(\psi, S),$$

respectively. Let the risk be the following sum:

$$\gamma(\psi, S) = \alpha(\psi) + \beta(\psi, S), \qquad \gamma_{nm,a}(\psi) = \sup_{S \in \mathcal{S}_{nm,a}} \gamma(\psi, S) = \alpha(\psi) + \beta_{nm,a}(\psi).$$

We define the minimax risk at fixed level $\alpha \in (0, 1)$ as

$$\beta_{nm,a,\alpha} = \inf_{\psi:\alpha(\psi)\le\alpha} \beta_{nm,a}(\psi).$$

Similarly, let the minimax testing risk be

$$\gamma_{nm,a} = \inf_{\psi} \gamma_{nm,a}(\psi).$$

From now on, we assume in the asymptotics that $N \to \infty$, $M \to \infty$ and $n = n_{NM} \to \infty$, $m = m_{NM} \to \infty$. Other assumptions will be given later.

We suppose that $a > 0$ is unknown. The aim of this paper is to give asymptotically sharp boundaries for minimax testing risk. It means that, first, we are interested in the conditions on $a = a_{NM}$ which guarantee distinguishability, that is, the fact that $\gamma_{nm,a} \to 0$ and $\beta_{nm,a,\alpha} \to 0$ for any $\alpha \in (0, 1)$. We construct a testing procedure based on a linear statistic combined with a scan statistic. We prove the upper bounds of the minimax testing risk of this procedure. Second, we describe conditions on a for which we have indistinguishability, that is, the convergence $\gamma_{nm,a} \to 1$ and $\beta_{nm,a,\alpha} \to 1 - \alpha$ for any $\alpha \in (0, 1)$. These results are called the lower bounds. The two sets of conditions are complementary and match in rate and constant.

Often the sizes n, m of submatrix are unknown, but we know a set $\mathcal{K}_{NM}$ of couples of indices $(n, m) \in \{1, \dots, N\} \times \{1, \dots, M\}$ containing the true one. Then we consider the "adaptive" problem for the combined alternative $\mathcal{S}_{NM,\mathbf{a}} = \bigcup_{(n,m)\in\mathcal{K}_{NM}} \mathcal{S}_{nm,a_{nm}}$, which corresponds to a collection $\mathbf{a} = \{a_{nm}, (n, m) \in \mathcal{K}_{NM}\}$. The quantities $\beta_{NM,\mathbf{a},\alpha}$, $\gamma_{NM,\mathbf{a}}$ are defined in a similar way as above. We define a testing procedure and check that, if a_{nm} satisfies the conditions for distinguishability uniformly over the collection $\mathbf{a}$, the upper bounds still hold. The adaptive lower bounds hold as an easy consequence of the minimax lower bounds.

The problem of choosing a submatrix in a Gaussian random matrix has been previously studied by Sun and Nobel [11]. They were interested in maximal size submatrices of a matrix with increasing size in two setups. First, they consider the case when the average of the entries of the submatrix is larger than a given threshold and, second, when the entries are well-fitted by a two-way ANOVA matrix in the least-squares sense (i.e., the sum of squares of residuals is smaller than some given threshold).

The algorithm of choosing such submatrices was previously introduced in Shabalin *et al.* [10], who were also interested in finding large average submatrices. This problem is strongly motivated by the research of gene expression in microarray data. In these large matrices, it is necessary to recover biclusters, that is associations between sets of samples (rows) and sets of variables (columns). These associations together with clinical and biological information are "a first step in identifying disease subtypes and gene regulatory networks". Many other algorithms for biclustering are discussed and compared on real-data bases concerning breast and lung cancer studies.

Similar problems were considered in Addario-Berry *et al.* [1]. They use the same testing procedures for vectors of random variables, where the alternatives may have various combinatorial structures. In particular, they consider the example of detecting a clique of a certain size in a graph and they compute upper and lower bounds for the Bayesian test error. A bipartite graph of size (N, M) is a graph having edges only between the N vertices of one set to the M vertices of a second set. A biclique is a complete bipartite subgraph of size (n, m), that is, a subgraph where all n vertices from the first set are connected to the m vertices from the second set. We consider the problem of detecting a biclique. Our results are sharp minimax and adaptive to the size of the unknown biclique.

The plan of the paper is as follows. In Section 2.1, we give the test procedures. We state the conditions on the detection boundary a such that distinguishability is possible. Under mild additional assumptions, we give the conditions on a so that the alternative is indistinguishable from the null hypothesis.

In Section 2.2, we consider the adaptive setup where (n, m) is unknown but belongs to some collection of sequences $\mathcal{K}_{NM}$. We compute the adaptive rates of testing of a slightly modified test procedure.

In Section 3, we perform a numerical study of the procedures that attain the sharp upper bounds. In order to compute the scan statistic, a heuristic stochastic algorithm from Shabalin *et al.* [10] is used. The empirical detection boundary is very close to the one predicted by our results.

In Section 4, we give extensions of our results to Gaussian variables of unknown variance σ^2, to non-Gaussian matrices with distribution in an exponential family and to two-sided tests for Gaussian matrices, respectively.

We include in Section 4.4 comments to understand how our results compare to previously studied alternatives: subsets without structure and rectangular submatrices. The first case can be assimilated to detection of a sparse signal in vector observations of length $N \times M$, so the set of alternatives and the detection boundary are much larger than in our case. We summarize well-known results by Ingster [7], Ingster and Suslina [8] and Donoho and Jin [6]. The second case is the detection of rectangles in the large matrix (connected submatrices), which constitutes a set of alternatives smaller than ours. This case is studied in Arias-Castro *et al.* [4] and [2] for other geometric shapes of clusters. In order to be self-contained, we state and prove sharp upper and lower bounds, for the rectangular clusters.

Section 5 is mainly concerned with the proof of the lower bounds stated in Section 2.1.2. The Appendix contains the proofs of the other results of the paper.

2. Main results

We denote by $n = n_{NM}$, $m = m_{NM}$ and $a = a_{N,M}$.

Denote also $p = n/N$, $q = m/M$. From now on, we suppose that

$$N \to \infty, \qquad M \to \infty, \qquad n \to \infty, \qquad m \to \infty \qquad \text{such that } p \to 0, q \to 0. \tag{2.1}$$

For general sequences $\{u_n\}_{n\geq 1}$ and $\{v_n\}_{n\geq 1}$ of real numbers, such that $v_n > 0$ for n large enough, we say that the sequences are asymptotically equivalent, $u_n \sim v_n$, if $\lim_{n\to\infty} u_n/v_n = 1$. Moreover, we say that the sequences are asymptotically of the same order, $u_n \asymp v_n$, if there exists two constants $0 < c \leq C < \infty$ such that $c \leq \liminf_{n\to\infty} u_n/v_n$ and $\limsup_{n\to\infty} u_n/v_n \leq C$.

2.1. Known size of the submatrix

In a minimax setup, we suppose that for each N, M we know n and m.

Let us consider two test procedures, one based on a linear statistic ψ_H^{lin} and the other based on a scan statistic $\psi^{\max}$. The final test procedure ψ^* will reject as soon as at least one of them rejects the null hypothesis.

2.1.1. *Test procedure and its performance*

The first test procedure ψ_H^{lin} is based on the linear statistic

$$t_{\mathrm{lin}} = \frac{1}{\sqrt{NM}} \sum_{i,j} Y_{ij}, \qquad \psi_H^{\mathrm{lin}} = \mathbb{1}_{t_{\mathrm{lin}} > H}.$$

The second test $\psi^{\max}$ is based on the maximal sum over all submatrices. Put

$$Y_C = \frac{1}{\sqrt{nm}} \sum_{(i,j)\in C} Y_{ij}, \tag{2.2}$$

and

$$t_{\max} = \max_{C\in\mathcal{C}_{nm}} Y_C, \qquad \psi^{\max} = \mathbb{1}_{t_{\max} > T_{nm}}, \tag{2.3}$$

where $T_{nm} = \sqrt{2\log(G_{nm})}$, $G_{nm} = \#(\mathcal{C}_{nm}) = \binom{N}{n}\binom{M}{m}$. The computation of this statistic is discussed in Section 3.

The following theorem gives sufficient conditions for the detection boundary a such that distinguishability holds. The test procedure which attains these bounds is

$$\psi^* = \max\{\psi_H^{\mathrm{lin}}, \psi^{\max}\}$$

for properly chosen H.

Theorem 2.1 (Upper bounds). *Assume* (2.1) *and let a be such that at least one of the following conditions hold*

$$a^2 nmpq \to \infty \tag{2.4}$$

or

$$\liminf \frac{a^2 nm}{2(n\log(p^{-1}) + m\log(q^{-1}))} > 1. \tag{2.5}$$

Then ψ^* *with* $H \to \infty$ *and such that* $H \le ca\sqrt{nmpq}, c < 1$ *when* (2.4) *holds, satisfies* $\gamma_{nm,a}(\psi^*) \to 0$.

Proof is given in Appendix A.1.

Formally, the procedure has a simple structure. Nevertheless, there are difficulties for computation of the scan statistic in the matrix case. Indeed, in the vector case, it is enough to order increasingly all the elements and take the sum of the largest values. In the matrix case, we have no such simple ordering. We shall discuss in the numerical study below the empirical algorithm used to compute the scan statistic.

Let us also note that this procedure assumes that n and m are known. An adaptive version of the scan test will be given in the next section.

2.1.2. *Lower bounds*

In this section, we obtain matching lower bounds that apply to all tests under additional assumptions on the matrix and submatrix sizes. We discuss these assumptions after the theorem.

Theorem 2.2 (Lower bounds). *Assume* (2.1) *and*

$$\frac{\log\log(p^{-1})}{\log(q^{-1})} \to 0, \qquad \frac{\log\log(q^{-1})}{\log(p^{-1})} \to 0. \tag{2.6}$$

Moreover, assume that

$$n\log(p^{-1}) \asymp m\log(q^{-1}), \tag{2.7}$$

and that the following two conditions are satisfied:

$$a^2 nmpq \to 0 \tag{2.8}$$

and

$$\limsup \frac{a^2 nm}{2(n\log(p^{-1}) + m\log(q^{-1}))} < 1. \tag{2.9}$$

Then the distinguishability is impossible, that is, $\gamma_{nm,a} \to 1$ *and* $\beta_{nm,a,\alpha} \to 1-\alpha$ *for any* $\alpha \in (0,1)$.

Proof is given in Section 5.

These results for the upper and the lower bounds can be interpreted as follows. Under the conditions (2.1), (2.6) and (2.7), a sharp detection boundary a^* is defined via the relations

$$(a^*)^2 nmpq \asymp 1, \qquad (a^*)^2 nm \sim 2(n\log(p^{-1}) + m\log(q^{-1})), \tag{2.10}$$

in the problem with known (n,m). Note that the detection boundary can be written as

$$a^* = \min\left\{\frac{1}{\sqrt{nmpq}}, \sqrt{\frac{2(n\log(p^{-1}) + m\log(q^{-1}))}{nm}}\right\}.$$

The additional assumptions (2.6) and (2.7) appearing in the previous lower bounds are satisfied, for example, in the case where $n \sim cm$, for some $0 < c < \infty$, and for $N \sim n^A$ and $M \sim m^B$ for A and B larger than 1. In this case, the detection boundary is of the form:

$$a^* \asymp n^{-2+(A+B)/2} \qquad \text{if } A+B \le 3,$$

$$a^* \sim \sqrt{\frac{2D\log(n)}{n}} \qquad \text{if } A+B > 3, \text{ where } D = (A-1)c + B - 1.$$

The particular case when $A = B > 1, c = 1$ is the case of asymptotically squared matrices and submatrices, and we get

$$a^* \asymp n^{-2+A} \qquad \text{if } A \le 3/2,$$
$$a^* \sim 2\sqrt{\frac{(A-1)\log(n)}{n}} \qquad \text{if } A > 3/2.$$

Remark 2.1. We can state the alternative hypothesis in a more general form:

$$H_1: \quad \exists C \in \mathcal{C}_{nm} \text{ such that } s_{ij} = 0 \text{ if } (i,j) \notin C \text{ and } \sum_{(i,j)\in C} s_{ij} \ge anm.$$

Indeed, our probabilities of error depend on the elements of the submatrix C only through the sum of its elements. Therefore, the previous test procedure will attain the same rates and the same lower bound techniques will give the previous results for this more general test problem.

2.2. Adaptation to the size of the submatrix

If the size (n, m) of the submatrix C with significantly large elements under the alternative (1.4) is unknown, we suppose that it belongs to the set $\mathcal{K}_{NM}$, for each N and M. The alternative hypothesis can be written

$$H_1(\mathcal{K}_{NM}): \quad \exists (n,m) \in \mathcal{K}_{NM}, \exists C \in \mathcal{C}_{nm} \text{ such that}$$
$$s_{ij} = 0, \text{ if } (i,j) \notin C \text{ and } s_{ij} \ge a_{nm}, \text{ if } (i,j) \in C.$$

Additionally, we suppose that the sequence of sets $\{\mathcal{K}_{NM}\}_{N,M}$ is such that

$$\sup_{(n,m)\in\mathcal{K}_{NM}} \left(\frac{1}{n} + \frac{1}{m} + \frac{n}{N} + \frac{m}{M}\right) \to 0$$

as $N, M \to \infty$.

This implies that

$$\sup_{(n,m)\in\mathcal{K}_{NM}} \left(\frac{\log(N)}{n\log(p^{-1})} + \frac{\log(M)}{m\log(q^{-1})}\right) \to 0 \qquad \text{as } N, M \to \infty. \tag{2.11}$$

The set $\mathcal{K}_{NM}$ contains sizes of submatrices that we have to explore in order to test in an adaptive way. Therefore, previous assumption insure, on the one hand, that $p \to 0$ and $q \to 0$ uniformly over $(n, m) \in \mathcal{K}_{NM}$ as $N, M \to \infty$ and, on the other hand, that the least size of the submatrices still grows to infinity with N and M.

The adaptive test procedure is $\psi^*_{NM} = \max\{\psi^{\text{lin}}_H, \psi^{\max}_{NM}\}$, where ψ^{lin}_H is the linear statistic defined in Section 2.1 and $\psi^{\max}_{NM}$ is a modified version of $\psi^{\max}$ defined as follows. Indeed, the linear statistic is free of n and m, but the scan statistic is not and, therefore, normalization will occur for each possible (n, m). Set

$$V_{nm} = \sqrt{2\log(NMG_{nm})}, \qquad t_{NM,\max} = \max_{(n,m)\in\mathcal{K}_{NM}} \max_{C\in\mathcal{C}_{nm}} Y_C/V_{nm}, \qquad \psi_{NM}^{\max} = \mathbb{1}_{t_{NM,\max}>1}.$$

The adaptive test will reject the null hypothesis as soon as at least one between the linear test or the scan tests associated to each $(n, m) \in \mathcal{K}_{NM}$ rejects.

Theorem 2.3. *Assume* (2.1) *and let the set* $\mathcal{K}_{NM}$ *be such that condition* (2.11) *holds.*

Upper bounds. *Let* $\mathbf{a} = \mathbf{a}_{NM} = \{a_{nm}, (n, m) \in \mathcal{K}_{NM}\}$ *be detection boundaries such that at least one of the following conditions hold*

$$\min_{(n,m)\in\mathcal{K}_{NM}} a_{nm}^2 nmpq \to \infty \tag{2.12}$$

or

$$\liminf \min_{(n,m)\in\mathcal{K}_{NM}} \frac{a_{nm}^2 nm}{2(n\log(p^{-1}) + m\log(q^{-1}))} > 1. \tag{2.13}$$

Then, ψ_{NM}^* *with* $H \to \infty$ *such that* $H \le c\min_{(n,m)\in\mathcal{K}_{NM}} a_{nm}\sqrt{nmpq}$ *for some* $0 < c < 1$ *when* (2.12) *holds, is such that* $\gamma_{NM,\mathbf{a}}(\psi_{NM}^*) \to 0$.

Proof is given in Appendix A.2.

The previous theorem actually shows that the test procedure is adaptive to the size (n, m) of the submatrix as far as the assumptions hold uniformly. Indeed, the linear procedure is free of the size of the submatrix and the scan statistic adapts to (n, m) without any loss in the rate.

The lower bounds in the adaptive setup are an obvious consequence of Theorem 2.2. Let us state the adaptive lower bounds: Suppose that for each N, M there exists (n^*, m^*) in the collection $\mathcal{K}_{NM}$ such that

$$\frac{\log\log(N/n^*)}{\log(M/m^*)} \to 0, \qquad \frac{\log\log(M/m^*)}{\log(N/n^*)} \to 0$$

and that $n^*\log(N/n^*) \asymp m^*\log(M/m^*)$, as $N \to \infty$ and $M \to \infty$. Let $\mathbf{a} = \mathbf{a}_{NM} = \{a_{nm}, (n, m) \in \mathcal{K}_{NM}\}$ be such that

$$a_{n^*m^*}^2 n^*m^*p^*q^* \to 0$$

and

$$\limsup \frac{a_{n^*m^*}^2 n^*m^*}{2(n^*\log(p^{*-1}) + m^*\log(q^{*-1}))} < 1.$$

Then $\gamma_{NM,\mathbf{a}} \to 1$ and $\beta_{NM,\mathbf{a},\alpha} \to 1 - \alpha$ for any $\alpha \in (0, 1)$.

3. Simulations

We have implemented the testing procedure $\psi^* = \max\{\psi_H^{\mathrm{lin}}, \psi^{\max}\}$ on synthetic data. While the linear procedure is rather obvious, the computation of the statistic $t_{\max} = \max_{C \in \mathcal{C}_{nm}} Y_C$ is done by using the heuristic algorithm introduced and studied empirically by Shabalin *et al.* [10]. This algorithm is also implemented and studied by Sun and Nobel [11] with good empirical results.

Let us briefly recall this algorithm: we choose randomly a set of n rows out of N. Then, we sum in every column the elements of the previously selected rows. We select now the columns corresponding to the m largest sums obtained in this way. We sum, next, in every row the elements belonging to the selected columns and select the rows corresponding to the n largest sums. We repeat the algorithm until the sum of elements Y_{ij} of the selected submatrix does not increase anymore. As the procedure can stop at a local maximum, we repeat the procedure K times, where K is large (in our simulation $K = 10\,000$). We take the maximum value of the outputs. This replication is needed to enforce that with high probability the output approaches the global maximum.

We have simulated matrices of size $N \times M$ of i.i.d. standard Gaussian random variables for $N = M = 200$ and $N = M = 500$.

We calibrated the test statistics ψ_H^{lin} and $\psi^{\max}$ in such a way that the type-I error occurs with probability $\alpha(\psi^*) \le 1\%$. This calibration is done by using the Gaussian quantile $H = 2.3262$ for ψ_H^{lin} and the empirical quantile (out of 100 samples) for $\psi^{\max}$.

Then, we have added the value $a > 0$ to the elements of the upper left submatrix of size $n \times m$. From resulting observations, we compute $\psi^* = \max\{\psi_H^{\mathrm{lin}}, \psi^{\max}\}$. We repeat the test $L = 100$ times and average the values of the test procedure ψ^*. Denote by $\bar{\psi}^*$ this average and note that $1 - \bar{\psi}^*$ estimates the probability of type-II error.

We plot the estimated second-type error probabilities for different values of a in the neighborhood of the detection boundary predicted by our theorems, for different values of n and m. The results in Figure 1 correspond to $N = M = 200$, while in Figure 2 to $N = M = 500$.

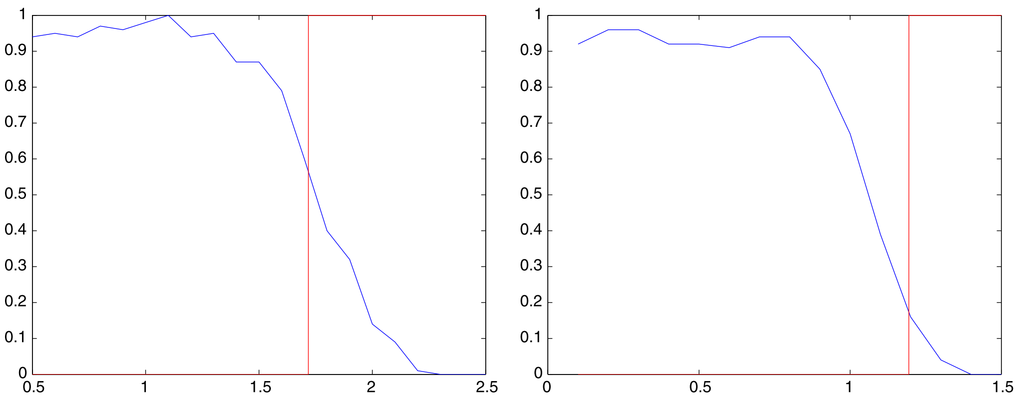

Figure 1. Estimated second-type error probability for fixed $\alpha = 1\%$, detection boundary a^*, $N = M = 200$; for $n = m = 5$, $a^* = 1.7179$ (left), for $n = m = 10$, $a^* = 1.1943$ (right).

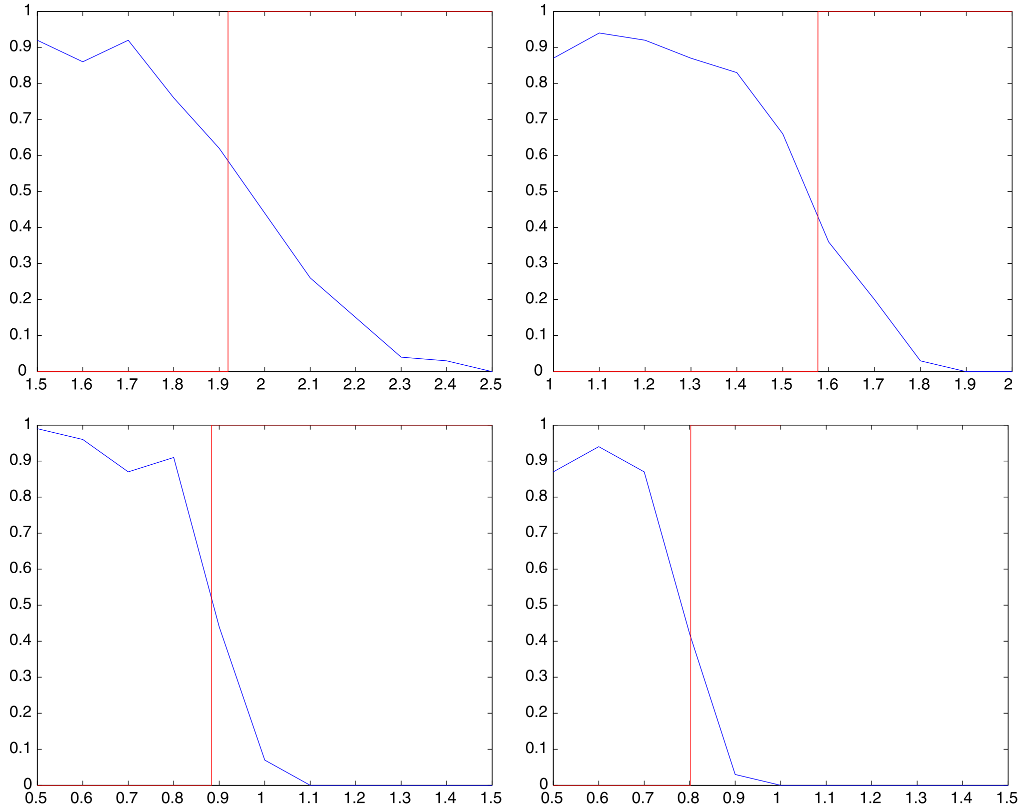

Figure 2. Estimated second-type error probability for fixed $\alpha = 1\%$, detection boundary a^*, $N = M = 500$; for $n = m = 5$, $a^* = 1.9194$ (upper left), for $n = 5$, $m = 10$, $a^* = 1.5767$ (upper right), for $n = 15$, $m = 20$, $a^* = 0.8831$ (lower left), for $n = m = 20$, $a^* = 0.8024$ (lower right).

Figures 1 and 2 show that the empirical detection boundary is very close to a^* which is predicted by out theoretical results. Indeed, the second-type error probability is close to 0.5 at some point close to a^*. The plots also show very fast decay of this probability on a small vicinity of a^*. This means that the test is very powerful for values of a slightly larger than the detection boundary a^*. Note also that, for fixed N and M, a^* decreases to 0 as n and m increase.

4. Extensions

We extend our results in different directions. First, we consider matrices of i.i.d. random variables having Gaussian law with unknown variance σ, next, random variables having a distribution belonging to the exponential family (not necessarily Gaussian) and, finally, test problem with two-sided alternative for the Gaussian matrices.

4.1. Extension to Gaussian variables with unknown variance

Sharp results in Theorems 2.1 and 2.2 still hold if the random variables Y_{ij} have unknown variance σ, under a mild additional assumption. We sketch here the test procedure and proof of the upper bounds.

We estimate the unknown variance σ^2 of our data by $\hat{\sigma}^2$, where

$$\hat{\sigma}^2 = \frac{1}{NM} \sum_{i,j} Y_{ij}^2.$$

This estimator is unbiased under the null hypothesis, but biased under the alternative.

We replace Y_{ij} by $Y_{ij}/\hat{\sigma}$ and slightly enlarge T_{nm} in the test procedure ψ^*. We denote by $\hat{t}_{\text{lin}} = t_{\text{lin}}/\hat{\sigma}$, $\hat{t}_{\max} = t_{\max}/\hat{\sigma}$ and put

$$\hat{\psi}^* = \max\{\mathbb{1}_{\hat{t}_{\text{lin}} > H}, \mathbb{1}_{\hat{t}_{\max} > T_{nm,\delta}}\}$$

for $T_{nm,\delta} = \sqrt{(2+\delta)\log(\binom{N}{n}\binom{M}{m})}$ and some $\delta > 0$ small enough. Recall that $T_{nm} = \sqrt{2\log(\binom{N}{n}\binom{M}{m})}$.

Theorem 4.1. *Assume* (2.1). *We suppose that alternatives under consideration are such that*

$$\tau_G := \frac{\max_{(i,j)\in C} s_{ij}}{\sum_{(i,j)\in C} s_{ij}} = \mathrm{o}(1). \tag{4.1}$$

If the quantity a is such that one of the following conditions hold

$$\frac{a}{\sigma}\sqrt{nmpq} \to \infty \quad \text{or} \quad \liminf \frac{a\sqrt{nm}}{\sqrt{2\sigma(n\log(p^{-1}) + m\log(q^{-1}))}} > 1 \tag{4.2}$$

then $\hat{\psi}^*$, *with* $H \to \infty$ *such that* $H^2 < \mathrm{o}(a^2 nmpq/\sigma^2 + \tau_G^{-1})$ *when* $a\sqrt{nmpq}/\sigma \to \infty$, *is such that* $\gamma_{nm,a}(\psi^*) \to 0$.

Proof is given in Appendix A.7.

Assumption (4.1) translates the fact that alternatives do not contain too "prominent" values s_{ij}. This holds when $\max_{(i,j)\in C} s_{ij} = \mathrm{o}(\sigma\sqrt{NM})$ and $a\sqrt{nmpq}/\sigma \to \infty$. The assumption ensures that the estimator of the unknown variance converges fast enough in order to detect the signal with the same rates as in the case of known variance. Moreover, one can slightly modify the proof and check that the condition (4.2) can be replaced by the condition $\max_{(i,j)\in C} s_{ij} = \mathrm{o}(\sigma\sqrt{NM})$ when $a\sqrt{nmpq}/\sigma = \mathrm{O}(1)$ as well.

4.2. Extension to general law from an exponential family

In many applications, we do not have Gaussian observations. Instead, we have observations X_{ij}, i.i.d. with probability density $g_{\theta_{ij}}$ from an exponential family, for all $i = 1, \dots, N$, and $j =$

$1, \dots, M$. We explain here how to use the previous testing procedures in order to deal with such setups and check that results similar to the case of Gaussian variables hold in this case. The exponential model will behave like a Gaussian model when the number of data is large, by asymptotic equivalence. We expect that the optimal detection boundary is the one for the Gaussian model properly rescaled.

We assume that the laws belong to an exponential family in the general form

$$g_\theta(x) = \mathrm{e}^{\eta(\theta)\cdot T(x) - C(\theta)} h(x), \qquad \theta \in \Theta \tag{4.3}$$

for the dominating measure μ, where η is supposed 2 times continuously differentiable and strictly increasing on Θ, that is, $\eta'(\theta) > 0$.

We consider a point θ^0 interior to Θ and test, based on X_{ij}'s, the null hypothesis H_0: $\theta_{ij} = \theta^0$ for all $i = 1, \dots, N$, $j = 1, \dots, M$, against the alternative

$$H_1: \quad \exists C \in \mathcal{C}_{nm} \text{ such that } \theta_{ij} = \theta^0 \text{ if } (i, j) \notin C \text{ and } \theta_{ij} - \theta^0 \geq d \text{ if } (i, j) \in C. \tag{4.4}$$

In order to build the test procedure as previously, we will rescale the observations as follows. First, put the exponential model in the canonical form, then change variables to $Y_{ij} = (T(X_{ij}) - m^0)/\sigma^0$, with $m^0 = E_{\theta^0}(T(X))$ and $\sigma^0 = \sqrt{\mathrm{Var}_{\theta^0}(T(X))}$ computed under the null hypothesis. Let us denote the common density of Y_{ij}'s by

$$f_s(y) = \mathrm{e}^{s\cdot y - A(s)} h(y),$$

where $s = \eta(\theta)\sigma_0$ and $A(s) = B(s/\sigma^0) - sm_0/\sigma_0$ and $B(\eta(\theta)) = C(\theta)$. Here, we have $A'(s^0) = 0$, $A''(s^0) = 1$ and

$$A\bigl(s^0 + a\bigr) - A\bigl(s^0\bigr) \sim \frac{a^2}{2} \quad \text{and} \quad A'\bigl(s^0 + a\bigr) \sim a \qquad \text{as } a \to 0. \tag{4.5}$$

In this way, the original problem corresponds to testing, based on Y_{ij}'s, the null hypothesis H_0: $s_{ij} = s^0$ for all $i = 1, \dots, N$, $j = 1, \dots, M$, against the alternative

$$H_1: \quad \exists C \in \mathcal{C}_{nm} \text{ such that } s_{ij} = s^0 \text{ if } (i, j) \notin C \text{ and } s_{ij} - s^0 \geq a \text{ if } (i, j) \in C.$$

We have the following results for exponential models.

Theorem 4.2. *Assume* (2.1). *We suppose that*

$$\frac{\log(p^{-1})}{m} + \frac{\log(q^{-1})}{n} \to 0. \tag{4.6}$$

Upper bounds. *If a is such that one of the following conditions hold*

$$A'\bigl(s^0 + a\bigr)\sqrt{nmpq} \to \infty \quad \text{or} \quad \liminf \frac{A'(s^0 + a)\sqrt{nm}}{\sqrt{2(n \log(p^{-1}) + m \log(q^{-1}))}} > 1$$

Table 1. Examples of calculations for testing in general exponential families

Probability law	η	m^0	σ^0	$\sqrt{I(\theta^0)} = \sigma^0\eta'(\theta^0)$
Poisson(θ), $\theta > 0$	$\log(\theta)$	θ^0	$\sqrt{\theta^0}$	$(\theta^0)^{-1/2}$
Ber(θ), $0 < \theta < 1$	$\log(\frac{\theta}{1-\theta})$	θ^0	$\sqrt{\theta^0(1-\theta^0)}$	$(\theta^0(1-\theta^0))^{-1/2}$
Exp(θ), $\theta > 0$	$-\theta^{-1}$	θ^0	θ^0	$(\theta^0)^{-1}$
$N(0,\theta^2)$, $\theta > 0$	$-\frac{1}{2\theta^2}$	$(\theta^0)^2$	$2(\theta^0)^2$	$2(\theta^0)^{-1}$

then ψ^*, *with* $H \to \infty$ *such that* $H \le cA'(s^0 + a)\sqrt{nmpq}$ *for some* $0 < c < 1$ *and with* T_{nm} *replaced by* $T_{nm,\delta}$ *for some* $\delta > 0$ *small enough, is such that* $\gamma_{nm,a}(\psi^*) \to 0$.

Lower bounds. *Assume, moreover, that conditions* (2.6) *and* (2.7) *hold. If a is such that the conditions* (2.8) *and* (2.9) *are satisfied, then* $\gamma_{nm,a} \to 1$ *and* $\beta_{nm,a,\alpha} \to 1-\alpha$ *for any* $\alpha \in (0,1)$.

Proof of the upper bounds is given in Appendix A.8.

The proof of the lower bounds uses the relation (4.5) and follows exactly the same lines as the proof of Theorem 2.2 in Section 5 except that we have to consider $T_{kl}^2 \sim (2+\delta)(k\log(p^{-1}) + l\log(q^{-1}))$ for some small $\delta > 0$ instead of thresholds in (5.3).

Under the assumption (4.6), the detection boundary $a^* \to 0$. Therefore,

$$A'\big(s^0 + a^*\big) \sim a^* \sim \big(\eta(\theta) - \eta\big(\theta^0\big)\big)\sigma^0 \sim \eta'\big(\theta^0\big)\sigma^0 d^*$$

as $d^* \to 0$. It is well known that the Fisher information at θ^0 in model (4.3) is $I(\theta_0) = (\sigma^0\eta'(\theta^0))^2$. In this way, we deduce the sharp asymptotic detection boundary for alternative (4.4) from Theorem 4.2: $d^* = a^*/\sqrt{I(\theta^0)}$.

Examples of such calculations for most popular probability distributions in the exponential family are given in Table 1.

4.3. Extension to two-sided alternative

Let us consider model (1.1) and the same null hypothesis (1.2), against the two-sided alternative:

$$H_1: \quad \exists C \in \mathcal{C}_{nm} \text{ such that } s_{ij} = 0 \text{ if } (i,j) \notin C \text{ and } |s_{ij}| \ge a \text{ if } (i,j) \in C.$$

Let us consider the following test procedures

$$z_{\text{lin}} = \frac{1}{\sqrt{2NM}} \sum_{i,j} \big(Y_{ij}^2 - 1\big) \quad \text{and} \quad \psi_{\text{lin}}^z = \mathbb{1}_{z_{\text{lin}} > H}$$

and

$$z_{\max} = \max_{C \in \mathcal{C}_{nm}} Z_C \qquad \text{where } Z_C = \frac{1}{\sqrt{2nm}} \sum_{(i,j)\in C} \big(Y_{ij}^2 - 1\big) \quad \text{and} \quad \psi_{\max}^z = \mathbb{1}_{z_{\max} > T_{nm,\delta}}$$

for some $\delta > 0$ small enough.

Theorem 4.3. *Assume* (2.1). *We suppose that* (4.6) *holds.*

Upper bounds. *If a is such that one of the following conditions hold*

$$a^2\sqrt{nmpq} \to \infty \quad \text{or} \quad \liminf \frac{a^2\sqrt{nm}}{2\sqrt{n\log(p^{-1}) + m\log(q^{-1})}} > 1$$

then $\psi^z = \max\{\psi^z_{\text{lin}}, \psi^z_{\max}\}$ *with* $H \to \infty$ *such that* $H \le ca^2/2\sqrt{nmpq}$ *for some* $0 < c < 1$, *is such that* $\gamma_{nm,a}(\psi^z) \to 0$.

Lower bounds. *Assume*, *moreover*, *that conditions* (2.6) *and* (2.7) *hold*. *If a is such that the following two conditions are satisfied*:

$$a^2\sqrt{nmpq} \to 0 \quad \text{and} \quad \limsup \frac{a^2\sqrt{nm}}{2\sqrt{n\log(p^{-1}) + m\log(q^{-1})}} < 1,$$

then $\gamma_{nm,a} \to 1$ *and* $\beta_{nm,a,\alpha} \to 1 - \alpha$ *for any* $\alpha \in (0, 1)$.

Proof is given in Appendix A.9.

4.4. Related testing problems

Let us consider again the model (1.1) and the null hypothesis (1.2). We shall see how our alternative which locates signal in submatrices of the large matrix compares to other alternatives. We consider first the alternatives where the signal is located anywhere (no structure: larger alternative) and then where the signal is located in block-submatrices (smaller alternative).

4.4.1. *Subsets without structure*

Let $\mathcal{D}_k$ consists of all subsets $D \subset \{1, \dots, N\} \times \{1, \dots, M\}$ of cardinality $\#(D) = k$ and let $k = nm$. Let us consider the alternative

$$H_1: \quad \exists D \in \mathcal{D}_{nm} \text{ such that } s_{ij} = 0 \text{ if } (i, j) \notin D \text{ and } s_{ij} \ge a \text{ if } (i, j) \in D \tag{4.7}$$

(we do not suppose that the set D is of product structure). Clearly, we can consider the matrix $\{Y_{ij}\}$ as a vector of dimension $P = NM$, and the problem is well studied as $P \to \infty$, see Ingster [7], Ingster and Suslina [8], Donoho and Jin [6].

The results are as follows. Let $k = P^{1-\beta}$, $\beta \in (0, 1)$. First, let $\beta \le 1/2$ which corresponds to $k^2 = O(P)$, that is, $(nm)^2 = O(NM)$. Then the detection boundary is determined by the first condition in (2.10). It means that distinguishability is impossible when $a^2 nmpq \to 0$. On the other hand, if $a^2 nmpq \to \infty$, then distinguishability is provided by the tests of the type ψ^{lin}_H.

Let $\beta \in (1/2, 1)$. Then the detection boundary is determined by the relation

$$a^* \sim \varphi(\beta)\sqrt{\log(P)} = \varphi(\beta)\sqrt{\log(NM)},$$

where

$$\varphi(\beta) = \begin{cases} \sqrt{2\beta - 1}, & 1/2 < \beta \le 3/4, \\ \sqrt{2}(1 - \sqrt{1-\beta}), & 3/4 < \beta < 1, \end{cases} \qquad \beta = 1 - \frac{\log(nm)}{\log(NM)}.$$

This means that, if $\limsup a/(\varphi(\beta)\sqrt{\log(NM)}) < 1$, then distinguishability is impossible, and if $\liminf a/(\varphi(\beta)\sqrt{\log(NM)}) > 1$, then distinguishability is provided by the "high criticism" tests $\psi^{\mathrm{HC}} = \mathbb{1}_{\{L_{\mathrm{HC}} > H\}}$ based on statistics

$$L(t) = \frac{\sum_{i,j}(\mathbb{1}_{\{Y_{ij} > t\}} - \Phi(-t))}{\sqrt{NM\Phi(t)\Phi(-t)}}, \qquad L_{\mathrm{HC}} = \max_{t_0 < t} L(t), \qquad t_0 > 0,$$

with $H = \sqrt{c \log\log(NM)}$, $c > 2$.

4.4.2. *Block-structured submatrices*

Let $\mathcal{E}_{nm}$ consist of all rectangles of size $n \times m$, that is, of the sets $E_{kl} = \{k+1, \dots, k+n\} \times \{l+1, \dots, l+m\}$, $0 \le k \le N-n$, $0 \le l \le M-m$, and the alternative is of the form

$$H_1\colon \quad \exists E \in \mathcal{E}_{nm} \text{ such that } s_{ij} = 0 \text{ if } (i,j) \notin E \text{ and } s_{ij} \ge a \text{ if } (i,j) \in E. \tag{4.8}$$

Similar problems were studied recently in Arias-Castro *et al.* [4] and [2] for other related geometrically-shaped clusters. Note that Arias-Castro *et al.* [4] also deals with detection of rectangular shapes in a square matrix.

The detection boundary for (4.8) is determined by

$$a^* \sim \sqrt{\frac{2(\log(p^{-1}) + \log(q^{-1}))}{nm}}.$$

Let us consider the test ψ_Z based on the scan statistic over a particular set of possible rectangles, which is a suitable "grid" on $\mathcal{E}_{nm}$ constructed as follows.

Take $\eta_{nm} = \eta > 0$. Put $n_k = (k-1)n\eta, k = 1, \dots, K, m_l = (l-1)m\eta, l = 1, \dots, L$, where K, L are such that $N - n(1+\eta) \le n_K \le N - n$, $M - m(1+\eta) \le m_L \le M - m$, which yield $K \sim N/(\eta n)$, $L \sim M/(\eta m)$. Put

$$Z_{kl} = \frac{1}{\sqrt{nm}} \sum_{(i,j) \in E_{n_k m_l}} Y_{ij}, \qquad Z = \max_{1 \le k \le K, 1 \le l \le L} Z_{kl}, \qquad \psi_Z = \mathbb{1}_{Z > \sqrt{2\log(KL)}}.$$

In this construction, we scan over a number $K \times L$ of rectangles which is much smaller than the cardinality of $\mathcal{E}_{nm}$ (for technical reasons) and which is also much larger than the set of non-overlapping rectangles (this set would not be large enough).

Theorem 4.4. *Assume* (2.1). *Then*

Upper bounds. *Let*

$$\liminf \frac{a^2 nm}{2(\log(p^{-1}) + \log(q^{-1}))} > 1,$$

Table 2. Table of sharp asymptotic rates of the detection boundary a^* for squared matrices and $n = N^{1-\beta}$

<table>
<tr><th>Rates</th><th>No structure (4.7)</th><th>Submatrix (1.4)</th><th>Block structure (4.8)</th></tr>
<tr><td>$\beta \in (0, \frac{1}{3}]$</td><td rowspan="2">$N^{-(1-2\beta)}$</td><td>$N^{-(1-2\beta)}$</td><td rowspan="3">$N^{-(1-\beta)}\sqrt{4\beta\log(N)}$</td></tr>
<tr><td>$\beta \in (\frac{1}{3}, \frac{1}{2}]$</td><td rowspan="2">$N^{-(1-\beta)/2}\sqrt{4\beta\log(N)}$</td></tr>
<tr><td>$\beta \in (\frac{1}{2}, 1)$</td><td>$\varphi(\beta)\sqrt{2\log(N)}$</td></tr>
</table>

and $\eta = \eta_{nm}$ *is taken in such way that* $\eta \to 0$, $n\eta \to \infty$, $m\eta \to \infty$, $|\log(\eta)| = o(|\log(pq)|)$. *Then* $\gamma_{nm,a}(\psi_Z) \to 0$ *for the test procedure* ψ_Z *previously described.*

Lower bounds. *Let*

$$\limsup \frac{a^2 nm}{2(\log(p^{-1}) + \log(q^{-1}))} < 1.$$

Then $\gamma_{nm,a} \to 1$, $\beta_{nm,a,\alpha} \to 1 - \alpha$ *for any* $\alpha \in (0, 1)$.

Proof is given in Appendix A.10.

Note that, the separation rates, that is, the asymptotics of a that provide distinguishability for the alternative (1.4), are intermediate between the fast separation rates for the alternative (4.8) and the slow rates for the alternative without structure (4.7).

Let us consider the particular case of squared matrices ($N = M$) and squared submatrices ($n = m$) such that $n = N^{1-\beta}$ for some $\beta \in (0, 1)$. The sharp asymptotic rates of the detection boundaries can be compared in Table 2.

5. Proof of Theorem 2.2

In the first part, we give the proof of the theorem and the other parts of this section are dedicated to proofs of intermediate results. More lemmas are in the Appendix.

We prove the lower bounds by first reducing the minimax testing error to a Bayesian testing risk with uniform prior over the set of parameters. Typically, one studies the likelihood ratio under the prior with respect to the law P_0 under the null hypothesis and proves that it tends to 1 in quadratic mean (under P_0). Nevertheless, this does not work as the covariance of the likelihood ratio is too large. Therefore, we truncate the likelihood ratio in a convenient way.

5.1. Prior and truncated likelihood ratio

Let $S_C = \{s_{ij}\}$ be the matrix such that $s_{ij} = 0$, $(i, j) \notin C$, $s_{ij} = a$, $(i, j) \in C$. Let us consider the prior on the set of matrices:

$$\pi = G_{nm}^{-1} \sum_{C \in \mathcal{C}_{nm}} \delta_{S_C}, \qquad G_{nm} = \#(\mathcal{C}_{nm}),$$

and let P_π be the mixture of likelihoods $P_\pi = G_{nm}^{-1}\sum_{C\in\mathcal{C}_{nm}} P_{S_C}$. Let us consider the likelihood ratio

$$L_\pi(Y) = \frac{\mathrm{d}P_\pi}{\mathrm{d}P_0}(Y) = G_{nm}^{-1}\sum_{C\in\mathcal{C}_{nm}} \frac{\mathrm{d}P_{S_C}}{\mathrm{d}P_0}(Y) = G_{nm}^{-1}\sum_{C\in\mathcal{C}_{nm}} \exp\bigl(-b^2/2 + bY_C\bigr),$$

here and below we set $b^2 \stackrel{\Delta}{=} a^2 nm$, and, for submatrix C of the size $n\times m$, the statistics Y_C are defined by (2.2). Since $\pi(\mathcal{S}_{nm}) = 1$, in order to obtain indistinguishability: $\gamma_{nm,a}\to 1$, $\beta_{nm,a,\alpha}\to 1-\alpha$, $\forall\alpha\in(0,1)$, it suffices to show

$$L_\pi(Y)\to 1 \qquad \text{in } P_0\text{-probability}. \tag{5.1}$$

Indeed,

$$\begin{aligned}
\gamma_{nm,a} &= \inf_{\psi\in[0,1]}\sup_{S\in\mathcal{S}_{nm,a}}\bigl(\alpha(\psi)+\beta(\psi,S)\bigr)\\
&\ge \inf_{\psi\in[0,1]}\frac{1}{G_{nm}}\sum_{S\in\mathcal{S}_{nm,a}}\left(E_0\bigl(\psi(Y)\bigr)+E_0\left[\bigl(1-\psi(Y)\bigr)\frac{\mathrm{d}P_{S_C}}{\mathrm{d}P_0}(Y)\right]\right)\\
&\ge \inf_{\psi\in[0,1]}\bigl(E_0\bigl(\psi(Y)\bigr)+E_0\bigl[\bigl(1-\psi(Y)\bigr)L_\pi(Y)\bigr]\bigr)\\
&\ge E_0\bigl(\psi^*(Y)\bigr)+E_0\bigl[\bigl(1-\psi^*(Y)\bigr)L_\pi(Y)\bigr],
\end{aligned}$$

where $\psi^*(Y) = \mathbb{1}_{L_\pi(Y)>1}$ is the likelihood ratio test. Therefore, (5.1) implies by Fatou's lemma that

$$\liminf\gamma_{nm,a}\ge E_0\bigl[\liminf\bigl(\psi^*(Y)+\bigl(1-\psi^*(Y)\bigr)L_\pi(Y)\bigr)\bigr],$$

that is, $\gamma_{nm,a}\to 1$. It is easy to deduce that $\beta_{nm,a,\alpha}\to 1-\alpha$.

Let us replace the statistics $L_\pi(Y)$ by their truncated version

$$\tilde{L}_\pi(Y) = G_{nm}^{-1}\sum_{C\in\mathcal{C}_{nm}}\frac{\mathrm{d}P_{S_C}}{\mathrm{d}P_0}(Y)\mathbb{1}_{\Gamma_C},$$

where the events Γ_C are determined as follows. Set

$$T_{kl} = \sqrt{2\bigl(\log(G_{kl})+\log(nm)\bigr)}\to\infty.$$

Take small $\delta_1>0$ (which will be specified later) and set $k_0=\delta_1 n$, $l_0=\delta_1 m$. Let $\mathcal{C}_{kl,C} = \{V\in\mathcal{C}_{kl}\colon\ V\subset C\}$ be the submatrices of $C\in\mathcal{C}_{nm}$ which are in $\mathcal{C}_{kl}$. Then we set

$$\Gamma_C = \bigcap_{k_0\le k\le n,\, l_0\le l\le m}\ \bigcap_{V\in\mathcal{C}_{kl,C}}\{Y_V\le T_{kl}\}. \tag{5.2}$$

By (A.1), under conditions on k,l in (5.2) (and similarly to the equivalent of T_{nm}^2) we have

$$T_{kl}^2\sim 2\bigl(k\log\bigl(p^{-1}\bigr)+l\log\bigl(q^{-1}\bigr)\bigr). \tag{5.3}$$

Indeed, when looking at second-order moments of the likelihood ratio $L_\pi(Y)$ a large contribution comes from overlapping submatrices C_1 and C_2 inducing correlated random variables Y_{C_1} and Y_{C_2}. Our idea is to truncate Y_V, for submatrices V of size close to (k,l), at its expected maximal value in order to reduce the contribution of these correlations.

Proposition 5.1. *Set* $\Gamma_{nm} = \bigcap_{C\in\mathcal{C}_{nm}} \Gamma_C$. *Then, under the assumptions of Theorem* 2.2, $P_0(\Gamma_{nm}) \to 1$.

Proof is given in Appendix A.3.

Proposition 5.1 yields

$$P_0\big(L_\pi(Y) = \tilde{L}_\pi(Y)\big) \to 1,$$

and in place of (5.1) it suffices to check that

$$\tilde{L}_\pi(Y) \to 1 \qquad \text{in } P_0\text{-probability}. \tag{5.4}$$

In order to get (5.4) it suffices to verify two relations:

Proposition 5.2. *Under the assumptions of Theorem* 2.2, *we have*

$$E_0(\tilde{L}_\pi) \to 1.$$

Proof is given in Appendix A.4.

Proposition 5.3. *Under the assumptions of Theorem* 2.2, *we have*

$$E_0\big(\tilde{L}_\pi^2\big) \le 1 + \mathrm{o}(1).$$

Propositions 5.2 and 5.3 imply that

$$E_0(\tilde{L}_\pi - 1)^2 = \big(E_0\big(\tilde{L}_\pi^2\big) - 1\big) - 2\big(E_0(\tilde{L}_\pi) - 1\big) \le \mathrm{o}(1),$$

which ends the proof of the theorem.

The remaining part of this section is devoted to obtaining the Proposition 5.3.

5.2. Proof of Proposition 5.3

We deal with the second order moment of the truncated likelihood ratio. We have

$$E_0\big(\tilde{L}_\pi^2\big) = G_{nm}^{-2} \sum_{C_1\in\mathcal{C}_{nm}, C_2\in\mathcal{C}_{nm}} E_0\big(\exp\big(-b^2 + b(Y_{C_1} + Y_{C_2})\big)\mathbb{1}_{\{\Gamma_{C_1}\cap\Gamma_{C_2}\}}\big). \tag{5.5}$$

We note that the expected value in the previous sum does not depend on C_1 and C_2 but merely on the size of their common submatrix. Let $C_1 = A_1 \times B_1$, $C_2 = A_2 \times B_2$ and set

$$k = \#(A_1 \cap A_2), \qquad l = \#(B_1 \cap B_2), \qquad V = (A_1 \cap A_2) \times (B_1 \cap B_2).$$

Under this notation, we put

$$g(k,l) = E_0\bigl(\exp\bigl(-b^2 + b(Y_{C_1} + Y_{C_2})\bigr)\mathbb{1}_{\{\Gamma_{C_1}\cap\Gamma_{C_2}\}}\bigr)$$

and see that we can rewrite (5.5) as follows:

$$E_0\bigl(\tilde{L}_\pi^2\bigr) = \sum_{k=0}^{n}\sum_{l=0}^{m} \frac{\#((C_1, C_2) \in \mathcal{C}_{nm}^2\colon \text{ size}(V) = (k,l))}{G_{nm}^2} g(k,l).$$

Notation. Set $z_{kl}^2 = a^2kl$, $\rho_{kl} = kl/nm$ and recall that $b^2 = a^2nm$. Note that it means also that $b^2 = z_{mn}^2$ and that $z_{kl}^2 = b^2\rho_{kl}$.

Lemma 5.1. *The following inequalities hold true.*

(1) *We have*

$$g(k,l) \le E_0\bigl(\exp\bigl(-b^2 + b(Y_{C_1} + Y_{C_2})\bigr)\bigr) = \exp\bigl(z_{kl}^2\bigr) \stackrel{\Delta}{=} g_1(k,l). \tag{5.6}$$

(2) *Let $b \ge T_{nm}/(1+\rho_{kl})$. Then*

$$\begin{aligned} g(k,l) &\le E_0\bigl(\exp\bigl(-b^2 + b(Y_{C_1} + Y_{C_2})\bigr)\mathbb{1}_{\{Y_{C_1}\le T_{nm}, Y_{C_2}\le T_{nm}\}}\bigr) \\ &\le \exp\left(-(T_{nm} - b)^2 + \frac{\rho_{kl}T_{nm}^2}{1+\rho_{kl}}\right) \stackrel{\Delta}{=} g_2(k,l). \end{aligned} \tag{5.7}$$

(3) *Let $k \ge \delta_1 n, l \ge \delta_1 m$, and $T_{kl} \le 2z_{kl}$. Then*

$$\begin{aligned} g(k,l) &\le E_0\bigl(\exp\bigl(-b^2 + b(Y_{C_1} + Y_{C_2})\bigr)\mathbb{1}_{\{Y_V\le T_{kl}\}}\bigr) \\ &= \exp\bigl(T_{kl}^2/2 - (T_{kl} - z_{kl})^2\bigr) \stackrel{\Delta}{=} g_3(k,l). \end{aligned} \tag{5.8}$$

Proof of Lemma 5.1 is given in Appendix A.5.

5.2.1. *From hypergeometric to binomial distributions*

Observe that the right-hand side of (5.5) is the expectation of $g(X_1, X_2)$ over X_1, X_2 which are independent and having hypergeometric distributions $\mathcal{HG}_1 = \mathcal{HG}(N,n,n), \mathcal{HG}_2 = \mathcal{HG}(M,m,m)$, respectively, that is,

$$E_0\bigl(\tilde{L}_\pi^2\bigr) = \sum_{k=0}^{n}\sum_{l=0}^{m} \frac{\binom{N}{n}\binom{n}{k}\binom{N-n}{n-k}\cdot\binom{M}{m}\binom{m}{l}\binom{M-m}{m-l}}{\binom{N}{n}^2\cdot\binom{M}{m}^2} g(k,l) = E_{\mathcal{HG}_1\times\mathcal{HG}_2} g(X_1, X_2). \tag{5.9}$$

Let us compare random variables X having hypergeometric distributions $\mathcal{HG} = \mathcal{HG}(N,n,n)$ and binomial distribution $\text{Bin} = \text{Bin}(n, \tilde{p})$, $\tilde{p} = n/(N-n)$.

Lemma 5.2. *Under binomial distribution, X is stochastically larger, than under hypergeometric distributions, that is, for any $x \in \mathbb{R}$,*

$$P_{\mathcal{HG}}(X \geq x) \leq P_{\text{Bin}}(X \geq x).$$

This yields, for any non-decreasing function g,

$$E_{\mathcal{HG}}\big(g(X)\big) \leq E_{\text{Bin}}\big(g(X)\big).$$

Proof. The first claim corresponds to Lemma 3 in Arias-Castro *et al.* [3]. The second claim follows from the Abel's transform of the series for the expectation. □

Let $P_{n,p}(k) = P_{\text{Bin}}(X = k)$, for some integer k, where X has binomial $\text{Bin}(n, p)$ distribution, and similarly $P_{N,n,n}(k) = P_{\mathcal{HG}}(X = k)$ for hypergeometric distributions $\mathcal{HG}(N, n, n)$ of X.

Lemma 5.3. *Let $n \to \infty$, $p \to 0$, $p > 0$, $k \geq n/r(p)$ where $r(p) \geq 1$ for $p > 0$ small enough, and $\log(r(p)) = \mathrm{o}(\log(p^{-1}))$. Then*

$$\log\big(P_{n,p}(k)\big) \leq k \log(p)\big(1 + \mathrm{o}(1)\big),$$
$$\log\big(P_{N,n,n}(k)\big) \leq k \log(p)\big(1 + \mathrm{o}(1)\big).$$

Proof is given in Appendix A.6.

5.2.2. *Evaluation of the expectation*

Take any small $\delta > 0$. The detection boundary a satisfies assumption (2.9), where the worst-case is when the limit is close to 1. It suffices, therefore, to consider the case

$$b^2 = a^2 nm \sim (2 - \delta)\big(n \log\big(p^{-1}\big) + m \log\big(q^{-1}\big)\big). \tag{5.10}$$

This implies

$$a^2 \asymp \frac{\log(p^{-1})}{m} + \frac{\log(q^{-1})}{n}. \tag{5.11}$$

In order to evaluate the right-hand side of (5.9), let us firstly divide the expectation into two parts $E_{\mathcal{HG}_1 \times \mathcal{HG}_2}(g(X_1, X_2)) = E_1 + E_2$, where

$$E_1 = E_{\mathcal{HG}_1 \times \mathcal{HG}_2}\big(g(X_1, X_2)\mathbb{1}_{X_1 a^2 < 1}\big), \qquad E_2 = E_{\mathcal{HG}_1 \times \mathcal{HG}_2}\big(g(X_1, X_2)\mathbb{1}_{X_1 a^2 \geq 1}\big).$$

Recall that we denote $\tilde{p} = n/(N - n)$, $\tilde{q} = m/(M - m)$ and the binomial distributions $\text{Bin}_1 = \text{Bin}(n, \tilde{p})$ and $\text{Bin}_2 = \text{Bin}(m, \tilde{q})$.

We would like to show that $E_1 \leq 1 + \mathrm{o}(1)$ and $E_2 = \mathrm{o}(1)$, so we keep in mind from now on that N, M are sufficiently large.

Evaluation of E_1.

It follows from (5.11) and (2.7) that $X_1 = \mathrm{O}(n/\log(q^{-1}))$ under $a^2 X_1 < 1$. By (5.6) we have

$$E_1 \le E_{\mathcal{HG}_1 \times \mathcal{HG}_2}\bigl(\exp\bigl(a^2 X_1 X_2\bigr)\mathbb{1}_{X_1 a^2 < 1}\bigr) \\ = E_{\mathcal{HG}_1}\bigl(E_{\mathcal{HG}_2}\bigl(\exp\bigl(a^2 X_1 X_2\bigr)\bigr)\mathbb{1}_{X_1 a^2 < 1}\bigr).$$

In view of Lemma 5.2, the expected value of a non-decreasing function of X_2 having hypergeometric distribution $\mathcal{HG}_2$ is less than the same expected value under the binomial $\mathrm{Bin}_2 = \mathrm{Bin}(m, \tilde{q}), \tilde{q} = m/(M-m)$,

$$E_{\mathcal{HG}_2}\bigl(\exp\bigl(a^2 X_1 X_2\bigr)\bigr) \\ \le E_{\mathrm{Bin}_2}\bigl(\exp\bigl(a^2 X_1 X_2\bigr)\bigr) = \bigl(1 + \tilde{q}\bigl(\mathrm{e}^{a^2 X_1} - 1\bigr)\bigr)^m \\ \le \exp\bigl(m\tilde{q}\bigl(\mathrm{e}^{a^2 X_1} - 1\bigr)\bigr).$$

Observe that, for some $B > 0$ under constraint $X_1 a^2 \le 1$,

$$\exp\bigl(m\tilde{q}\bigl(\mathrm{e}^{a^2 X_1} - 1\bigr)\bigr) \le \exp\bigl(Bmqa^2 X_1\bigr).$$

Taking the expectation over X_1, we get similarly

$$E_1 \le E_{\mathcal{HG}_1}\bigl(\exp\bigl(Bmqa^2 X_1\bigr)\bigr) \\ \le E_{\mathrm{Bin}_1}\bigl(\exp\bigl(Bmqa^2 X_1\bigr)\bigr) = \bigl(1 + \tilde{p}\bigl(\mathrm{e}^{Bm\tilde{q}a^2} - 1\bigr)\bigr)^n \\ \le \exp\bigl(n\tilde{p}\bigl(\mathrm{e}^{Bmqa^2} - 1\bigr)\bigr).$$

By (5.11) and (2.7), we have $a^2 m \asymp \log(p^{-1})$. By condition (2.6), we have $q \log(p^{-1}) = \mathrm{o}(1)$, which yields $mqa^2 = \mathrm{o}(1)$. Thus, $n\tilde{p}(\mathrm{e}^{Bmqa^2} - 1) \sim Bnpmqa^2 = \mathrm{o}(1)$ by the condition (2.8), and we get

$$E_1 \le \exp\bigl(\mathrm{o}(1)\bigr) = 1 + \mathrm{o}(1).$$

Evaluation of E_2.

In order to evaluate E_2, we use the assumption (2.7) and write

$$a^2 \asymp \log\bigl(p^{-1}\bigr)/m \asymp \log\bigl(q^{-1}\bigr)/n$$

instead of (5.11). Therefore, we can take $\delta_1 > 0$ small enough such that $\delta_1 a^2 m \le \log(p^{-1})/2$ and $\delta_1 a^2 n \le \log(q^{-1})/2$ for N, M, n and m large enough.

Divide E_2 into two parts $E_2 = E_{21} + E_{22}$, where

$$E_{21} = E_{\mathcal{HG}_1 \times \mathcal{HG}_2}\bigl(g(X_1, X_2)\mathbb{1}_{X_1 a^2 \ge 1, X_2/m < \delta_1}\bigr), \\ E_{22} = E_{\mathcal{HG}_1 \times \mathcal{HG}_2}\bigl(g(X_1, X_2)\mathbb{1}_{X_1 a^2 \ge 1, X_2/m \ge \delta_1}\bigr).$$

Evaluation of E_{21}.

By applying Lemma 5.3 with $\log r(p) \overset{\Delta}{=} \log(\log(q^{-1}))$, which is $\mathrm{o}(\log(p^{-1}))$ by (2.6), and since $P_{M,m,m}(l) \le 1$, we get

$$E_{21} \le \sum_{n \ge k > a^{-2}, 0 \le l \le \delta_1 m} \exp\bigl(a^2 kl\bigr) P_{N,n,n}(k) P_{M,m,m}(l)$$
$$\le \sum_{n \ge k > a^{-2}, 0 \le l \le \delta_1 m} \exp\bigl(k\bigl(a^2 l - \log\bigl(p^{-1}\bigr)\bigl(1 + \mathrm{o}(1)\bigr)\bigr)\bigr).$$

Observe that under the constraints in the sum,

$$a^2 l \le \delta_1 a^2 m \le \log\bigl(p^{-1}\bigr)\bigl(1/2 + \mathrm{o}(1)\bigr),$$

which yields in the previous exponential

$$k\bigl(a^2 l - \log\bigl(p^{-1}\bigr)\bigl(1 + \mathrm{o}(1)\bigr)\bigr) \le -k \log\bigl(p^{-1}\bigr)\bigl(1/2 + \mathrm{o}(1)\bigr)$$
$$\le -a^{-2} \log\bigl(p^{-1}\bigr)\bigl(1/2 + \mathrm{o}(1)\bigr) \asymp m.$$

Therefore, we have $E_{21} \le nm \exp(-Bm) = \mathrm{o}(1)$ for some $B > 0$ by condition (2.6).

Evaluation of E_{22}.

In order to evaluate the item E_{22}, we divide it in two parts as well: $E_{22} = I_1 + I_2$,

$$I_1 = E_{\mathcal{H}\mathcal{G}_1 \times \mathcal{H}\mathcal{G}_2}\bigl(g(X_1, X_2) \mathbb{1}_{X_1 a^2 \ge 1, X_1/n < \delta_1, X_2/m \ge \delta_1}\bigr),$$
$$I_2 = E_{\mathcal{H}\mathcal{G}_1 \times \mathcal{H}\mathcal{G}_2}\bigl(g(X_1, X_2) \mathbb{1}_{X_1/n \ge \delta_1, X_2/m \ge \delta_1}\bigr).$$

The evaluation of I_1 is similar to the evaluation of E_{21} and we get $I_1 = \mathrm{o}(1)$.

Evaluation of I_2.

Let us divide the set $\mathcal{H} = \{(k,l)\colon k/n \ge \delta_1, l/m \ge \delta_1\}$, appearing in I_2, into two parts:

$$\mathcal{H}_1 = \bigl\{(k,l) \in \mathcal{H}\colon k \log\bigl(p^{-1}\bigr) + l \log\bigl(q^{-1}\bigr) \ge 2\rho_{kl}\bigl(n \log\bigl(p^{-1}\bigr) + m \log\bigl(q^{-1}\bigr)\bigr)\bigr\},$$
$$\mathcal{H}_2 = \bigl\{(k,l) \in \mathcal{H}\colon k \log\bigl(p^{-1}\bigr) + l \log\bigl(q^{-1}\bigr) < 2\rho_{kl}\bigl(n \log\bigl(p^{-1}\bigr) + m \log\bigl(q^{-1}\bigr)\bigr)\bigr\}.$$

This yields the division of I_2 into $I_2 = I_{12} + I_{22}$. Observe that $\rho_{kl} \ge \delta_1^2$ for $(k,l) \in \mathcal{H}$.

Let us consider I_{12}. Recalling (5.7), observe that we can take $\delta > 0$ small enough in (5.10) such that $t = T_{nm} - b(1 + \rho_{kl}) < 0$. Applying (5.7) and Lemma 5.3 for $P_{N,n,n}(k)$, $P_{M,m,m}(l)$, we get

$$I_{12} \le \sum_{(k,l) \in \mathcal{H}_1} \exp\biggl(-(T_{nm} - b)^2 + \frac{T_{nm}^2 \rho_{kl}}{1 + \rho_{kl}} - k \log\bigl(p^{-1}\bigr) - l \log\bigl(q^{-1}\bigr) + \mathrm{o}\bigl(T_{nm}^2\bigr)\biggr).$$

Note that for $\delta > 0$ small enough in (5.10) one can take $\delta_2 = \delta_2(\delta) > 0$ such that $(T_{nm} - b)^2 \ge \delta_2 T_{nm}^2$ for the first item in the exponent. Denote

$$A = A_{n,p} = n \log\bigl(p^{-1}\bigr); \qquad B = B_{m,q} = m \log\bigl(q^{-1}\bigr)$$

and $T_{nm}^2 = 2\log(G_{nm}) \sim 2(A+B)$ by (A.1). Observe that $2(A+B)\rho_{kl} \le A\frac{k}{n} + B\frac{l}{m}$ for $(k,l) \in \mathcal{H}_1$.

The following terms in the power of the exponential above can be bounded on the set $\mathcal{H}_1$ as follows

$$\begin{aligned}\frac{T_{nm}^2\rho_{kl}}{1+\rho_{kl}} - k\log(p^{-1}) - l\log(q^{-1}) &= \frac{2(A+B)\rho_{kl}}{1+\rho_{kl}} - \left(A\frac{k}{n} + B\frac{l}{m}\right) + \mathrm{o}(T_{nm}^2)\\ &\le \left(\frac{1}{1+\rho_{kl}} - 1\right)\left(A\frac{k}{n} + B\frac{l}{m}\right) + \mathrm{o}(T_{nm}^2) \le \mathrm{o}(T_{nm}^2).\end{aligned}$$

Therefore,

$$I_{12} \le 2nm\exp\bigl(-(\delta_2 + \mathrm{o}(1))T_{nm}^2\bigr) = \mathrm{o}(1).$$

Consider now the item I_{22}. Recalling (5.3), (5.10) and $z_{kl}^2 = \rho_{kl}a_{nm}^2$ observe that the constraint in $\mathcal{H}_2$ corresponds to $T_{kl}^2 < 2z_{kl}^2(1+\mathrm{o}(1))$. This implies $T_{kl}^2 < 4z_{kl}^2$ and $T_{kl} - 2z_{kl} < 0$ for N, M large enough.

Applying (5.8) and Lemma 5.3 for $P_{N,n,n}(k)$, $P_{M,m,m}(l)$, we similarly get

$$I_{22} \le \sum_{(k,l)\in\mathcal{H}_2} \exp\bigl(T_{kl}^2/2 - (T_{kl} - z_{kl})^2 - \bigl(k\log(p^{-1}) + l\log(q^{-1})\bigr)\bigl(1+\mathrm{o}(1)\bigr)\bigr).$$

Since $k\log(p^{-1}) + l\log(q^{-1}) \sim T_{kl}^2/2$, the power in the exponent is of the form

$$-(T_{kl} - z_{kl})^2 + \mathrm{o}(T_{kl}^2).$$

Under (5.10), we can see that $(T_{kl} - z_{kl})^2 \ge \delta_2 T_{kl}^2$ for some $\delta_2 > 0$ and N, M large enough. In fact, recalling $A > 0$, $B > 0$, $k/n \in (\delta_1, 1]$, $l/m \in (\delta_1, 1]$ before we have

$$\begin{aligned}T_{kl}^2 - z_{kl}^2 &= 2\left(A\frac{k}{n} + B\frac{l}{m}\right) - (2-\delta)\rho_{kl}(A+B) + \mathrm{o}(T_{kl}^2)\\ &= \delta(A+B)\rho_{kl} + 2A\frac{k}{n}\left(1 - \frac{k}{n}\right) + 2B\frac{l}{m}\left(1 - \frac{l}{m}\right) + \mathrm{o}(T_{kl}^2)\\ &\ge \delta(A+B)\rho_{kl} + \mathrm{o}(T_{kl}^2) \sim \delta z_{kl}^2.\end{aligned}$$

Since $T_{kl} \asymp T_{nm}$ for $(k,l) \in \mathcal{H}$, these yield $I_{22} = \mathrm{o}(1)$.

Proposition 5.3 follows.

Appendix

A.1. Proof of Theorem 2.1

It is easy to see that $\alpha(\psi^*) \le \alpha(\psi_H^{\mathrm{lin}}) + \alpha(\psi^{\max})$ and that $\beta_{nm,a}(\psi^*) \le \min\{\beta_{nm,a}(\psi_H^{\mathrm{lin}}), \beta_{nm,a}(\psi^{\max})\}$. Therefore, we study the two test procedures separately.

We have, for any real number H,

$$\alpha(\psi_H^{\text{lin}}) = \Phi(-H), \qquad \beta_{nm,a}(\psi_H^{\text{lin}}) \le \Phi(H - a\sqrt{nmpq}),$$

where Φ denotes the cumulative distribution function of a standard Gaussian random variable. Indeed, observe that the statistic t_{lin} is standard Gaussian under P_0 which yields the relation for $\alpha(\psi_H^{\text{lin}})$. Also $t_{\text{lin}} \sim \mathcal{N}(h_{S_C}, 1)$ under P_{S_C}-probability, $S_C \in \mathcal{S}_{nm,a}$, where

$$h_{S_C} = \frac{1}{\sqrt{NM}} \sum_{(i,j)\in C} s_{ij} \ge \frac{anm}{\sqrt{NM}} = a\sqrt{mnpq}.$$

This yields the relation $\beta(\psi_H^{\text{lin}}, S) \le \Phi(H - a\sqrt{mnpq}), \forall S \in \mathcal{S}_{nm,a}$ and the same inequality for $\beta_{nm,a}(\psi_H^{\text{lin}})$.

As a consequence, if $H \to \infty$ then $\alpha(\psi_H^{\text{lin}}) = \Phi(-H) \to 0$ and if (2.4) holds and $H \le ca\sqrt{nmpq}$,

$$\beta_{nm,a}(\psi_H^{\text{lin}}) \le \Phi\big((c-1)a\sqrt{nmpq}\big) \to 0$$

for $c < 1$. Thus $\gamma_{nm,a}(\psi_H^{\text{lin}}) \to 0$.

Now, we prove that, $\alpha(\psi^{\max}) \to 0$ and $\beta_{nm,a}(\psi^{\max}) \le \Phi(T_{nm} - a\sqrt{nm})$, under (2.1). It is not hard to check that under (2.1),

$$\log(G_{nm}) \sim \big(n\log(p^{-1}) + m\log(q^{-1})\big). \tag{A.1}$$

Observe that $Y_C \sim \mathcal{N}(0,1)$ under P_0 and, since $G_{nm} \to \infty$ and $\Phi(-T) \asymp \exp(-T^2/2)/T$ as $T \to \infty$, we get

$$\alpha(\psi^{\max}) = P_0(t_{\max} > T_{nm}) \le \sum_{C\in\mathcal{C}_{nm}} P_0(Y_C > T_{nm}) = G_{nm}\Phi(-T_{nm}) \to 0.$$

Let $S_C \in \mathcal{S}_{nm,a}$. Then $Y_C \sim \mathcal{N}(g_{S_C}, 1)$ under P_{S_C}-probability with

$$g_{S_C} = (nm)^{-1/2} \sum_{(i,j)\in C} s_{ij} \ge a\sqrt{nm}.$$

Observe that

$$\begin{aligned}\beta(\psi^{\max}, S_C) &= P_{S_C}(t_{\max} \le T_{nm}) \le P_{S_C}(Y_C \le T_{nm}) = \Phi(T_{nm} - g_{S_C}) \\ &\le \Phi(T_{nm} - a\sqrt{nm}).\end{aligned}$$

Thus, under (2.5),

$$a\sqrt{nm} - T_{nm} = T_{nm}\left(\frac{a\sqrt{nm}(1+\mathrm{o}(1))}{\sqrt{2(n\log(p^{-1}) + m\log(q^{-1}))}} - 1\right) \to \infty$$

and this implies that $\gamma_{nm,a}(\psi^{\max}) \to 0$.

A.2. Proof of Theorem 2.3

Note that the test ψ_H^{lin} does not depend on n, m. Therefore for distinguishability in the adaptive problem it is sufficient to assume (2.12) (which is a uniform version of (2.4)). We have $\gamma_{NM,\mathbf{a}}(\psi_H^{\mathrm{lin}}) \to 0$.

For the test $\psi_{NM}^{\max}$, we obtain similarly to the nonadaptive case, that

$$\alpha\bigl(\psi_{NM}^{\max}\bigr) = P_0(t_{NM,\max} > 1) \le \sum_{n,m} \sum_{C \in \mathcal{C}_{nm}} P_0(Y_C > H_{nm}) = \sum_{n,m} \sum_{C \in \mathcal{C}_{nm}} \Phi(-H_{nm})$$

$$\asymp \sum_{n,m} \sum_{C \in \mathcal{C}_{nm}} \frac{1}{NMG_{nm}\sqrt{\log(NMG_{nm})}} = \mathrm{O}\biggl(\frac{1}{\sqrt{\log(NM)}}\biggr) \to 0.$$

Moreover,

$$\beta\bigl(\psi_{NM}^{\max}, S_C\bigr) = P_{S_C}(t_{NM,\max} \le 1) \le P_{S_C}(Y_C \le V_{nm}) = \Phi(V_{nm} - g_{S_C})$$
$$\le \Phi(V_{nm} - a_{nm}\sqrt{nm})$$

and we deduce that

$$\beta_{nm,\mathbf{a}}\bigl(\psi_{NM}^{\max}\bigr) \le \Phi\Bigl(\max_{(n,m)\in\mathcal{K}_{NM}} (V_{nm} - a_{nm}\sqrt{nm})\Bigr).$$

By (A.1), we have

$$\min_{(n,m)\in\mathcal{K}_{NM}} (a_{nm}\sqrt{nm} - V_{nm})$$
$$= \min_{(n,m)\in\mathcal{K}_{NM}} T_{nm}\biggl(\frac{a_{nm}\sqrt{nm}(1+\mathrm{o}(1))}{\sqrt{2(n\log(p^{-1}) + m\log(q^{-1}))}} - \sqrt{1 + \frac{\log(NM)}{n\log(p^{-1}) + m\log(q^{-1})}}\biggr),$$

which goes to infinity under (2.11) and (2.13). Thus, we have $\gamma_{NM,\mathbf{a}}(\psi_{NM}^{\max}) \to 0$.

A.3. Proof of Proposition 5.1

It suffices to check that $P_0(\Gamma_{nm}^c) \to 0$, where A^c states for the complement of the event A. We have

$$\Gamma_{nm}^c = \bigcup_{C\in\mathcal{C}_{nm}} \bigcup_{k_0\le k\le n, l_0\le l\le m} \bigcup_{V\in\mathcal{C}_{kl,C}} \{Y_V > T_{kl}\}$$
$$= \bigcup_{k_0\le k\le n, l_0\le l\le m} \bigcup_{V\in\mathcal{C}_{kl}} \{Y_V > T_{kl}\}.$$

Since $Y_V \sim \mathcal{N}(0,1)$ under P_0, we have, by definition of T_{kl} and using the asymptotics $\Phi(-x) \sim \mathrm{e}^{-x^2/2}/\sqrt{2\pi}x$, $x \to \infty$,

$$P_0\bigl(\Gamma_{nm}^c\bigr) \le \sum_{k_0 \le k \le n, l_0 \le l \le m} \sum_{V \in \mathcal{C}_{kl}} \Phi(-T_{kl}) = \sum_{k_0 \le k \le n, l_0 \le l \le m} G_{kl}\Phi(-T_{kl})$$
$$\le \sum_{k_0 \le k \le n, l_0 \le l \le m} \frac{1+\mathrm{o}(1)}{nmT_{kl}\sqrt{2\pi}} \to 0.$$

Proposition 5.1 follows.

A.4. Proof of Proposition 5.2

In view of symmetry in C, it suffices to check that, for any fixed $C \in \mathcal{C}_{nm}$,

$$E_0\biggl(\frac{\mathrm{d}P_{S_C}}{\mathrm{d}P_0}\mathbb{1}_{\Gamma_C}\biggr) = P_{S_C}(\Gamma_C) \to 1,$$

or, equivalently, $P_{S_C}(\Gamma_C^c) \to 0$. Set $z_{kl}^2 = a^2kl$. Since $Y_V \sim \mathcal{N}(z_{kl},1)$ under P_{S_C} for $V \in \mathcal{C}_{kl,C}$, we have

$$P_{S_C}\bigl(\Gamma_C^c\bigr) \le \sum_{k_0 \le k \le n, l_0 \le l \le m} \sum_{V \in \mathcal{C}_{kl,C}} \Phi(z_{kl} - T_{kl}) = \sum_{k_0 \le k \le n, l_0 \le l \le m} G_{kl}^{mn}\Phi(z_{kl} - T_{kl}),$$

where $G_{kl}^{mn} = \#(\mathcal{C}_{kl,C}) = \binom{n}{k}\binom{m}{l}$. Under assumptions (2.6) and (2.9) there exists $\delta > 0$ such that

$$b^2 = a^2nm < (2-\delta)\bigl(n\log\bigl(p^{-1}\bigr) + m\log\bigl(q^{-1}\bigr)\bigr). \tag{A.2}$$

Let us show that under (A.2) one has $z_{kl}^2 < T_{kl}^2(1-\delta/2+\mathrm{o}(1))$. In fact, since $\delta_1 n \le k \le n$, $\delta_1 m \le l \le m$, and by (5.3), we have

$$\begin{aligned} z_{kl}^2 = a^2kl &\le (2-\delta)\bigl(k(l/m)\log\bigl(p^{-1}\bigr) + l(k/n)\log\bigl(q^{-1}\bigr)\bigr) \\ &\le (2-\delta)\bigl(k\log\bigl(p^{-1}\bigr) + l\log\bigl(q^{-1}\bigr)\bigr) \sim (1-\delta/2)T_{kl}^2. \end{aligned} \tag{A.3}$$

Thus we get, for some $\delta_2 > 0$,

$$\Phi(z_{kl} - T_{kl}) \le \exp\bigl(-\delta_2 T_{kl}^2\bigr).$$

Observe now that, under constraints on $\delta_1 n \le k \le n$, $\delta_1 m \le l \le m$ we have $\log(G_{kl}^{nm}) = \mathrm{O}(n + m)$. This follows from evaluations similar to the proof of (A.1). On the other hand, we have $T_{kl}^2 \sim 2(k\log(p^{-1}) + l\log(q^{-1})) \gg (n+m)$ under the same constraints. This yields

$$\sum_{k_0 \le k \le n, l_0 \le l \le m} G_{kl}^{mn}\Phi(z_{kl} - T_{kl}) \le \sum_{k_0 \le k \le n, l_0 \le l \le m} \exp\bigl(\mathrm{O}(n+m) - \delta_2 T_{kl}^2\bigr) \to 0.$$

Proposition 5.2 follows.

A.5. Proof of Lemma 5.1

The first inequalities in (5.6)–(5.8) are evident, and we will prove the second ones. The proofs are based on the well-known relation: if $X \sim \mathcal{N}(0,1)$, then

$$E\big(\exp(\tau X)\big) = \exp\big(\tau^2/2\big) \qquad \forall \tau \in \mathbb{R}. \tag{A.4}$$

Let $V_1 = C_1 \setminus C_2$, $V_2 = C_2 \setminus C_1$, $V = C_1 \cap C_2$, and observe that the sets V_1, V_2, V are disjoint, $C_1 = V_1 \cup V$, $C_2 = V_2 \cup V$ and $\#(V_1) = \#(V_2) = nm - kl$, $\#(V) = kl$.

Let $0 < kl < nm$. Let us write the statistics Y_{C_1}, Y_{C_2} in a more convenient form

$$Y_{C_1} = \sqrt{1-\rho_{kl}}\,Y_{V_1} + \sqrt{\rho_{kl}}\,Y_V, \qquad Y_{C_2} = \sqrt{1-\rho_{kl}}\,Y_{V_2} + \sqrt{\rho_{kl}}\,Y_V,$$

where as above, for $U \subset \{1,\dots,N\} \times \{1,\dots,M\}$, $\#(U) > 0$ we set

$$Y_U = \frac{1}{\sqrt{\#(U)}} \sum_{(i,j)\in U} Y_{ij}.$$

Observe that Y_{V_1}, Y_{V_2}, Y_V are standard Gaussian and independent under P_0.

Recall that $b = a\sqrt{nm}$ and put $c = b\sqrt{1-\rho_{kl}}$. It is obvious that $b^2 = c^2 + z_{kl}^2$. Moreover, by applying (A.4), we get

$$\begin{aligned}
&E_0\big(\exp\big(-b^2 + b(Y_{C_1} + Y_{C_2})\big)\big)\\
&\quad = E_0\big(\exp\big(-c^2/2 + cY_{V_1}\big)\big) \cdot E_0\big(\exp\big(-c^2/2 + cY_{V_2}\big)\big) \cdot E_0\big(\exp\big(-z_{kl}^2 + 2z_{kl}Y_V\big)\big)\\
&\quad = \exp\big(z_{kl}^2\big).
\end{aligned}$$

If $kl = 0$ or $kl = nm$, we can prove this in a similar way. Lemma 5.1 (5.6) follows.

In order to get the second inequality, observe that, for $0 < kl < nm$ and for any $h \geq 0$,

$$\begin{aligned}
&E_0\big(\exp\big(-b^2 + b(Y_{C_1} + Y_{C_2})\big)\mathbb{1}_{Y_{C_1}\leq T_{nm}, Y_{C_2}\leq T_{nm}}\big)\\
&\quad \leq \mathrm{e}^{-b^2+2T_{nm}h} E_0\big(\exp\big((b-h)(Y_{C_1}+Y_{C_2}) + h(Y_{C_1}+Y_{C_2}-2T_{nm})\big)\mathbb{1}_{Y_{C_1}\leq T_{nm}, Y_{C_2}\leq T_{nm}}\big)\\
&\quad \leq \mathrm{e}^{-b^2+2T_{nm}h} E_0\big(\exp\big((b-h)(Y_{C_1}+Y_{C_2})\big)\big)\\
&\quad = \mathrm{e}^{-b^2+2T_{nm}h} E_0\big(\exp\big((b-h)(1-\rho_{kl})^{1/2}(Y_{V_1}+Y_{V_2}) + 2(b-h)\rho_{kl}^{1/2}Y_V\big)\big)\\
&\quad = \exp\big(-b^2 + 2T_{nm}h + (b-h)^2(1-\rho_{kl}) + 2(b-h)^2\rho_{kl}\big)\\
&\quad = \exp\big(-b^2 + 2T_{nm}h + (b-h)^2(1+\rho_{kl})\big).
\end{aligned}$$

Taking $h = b - T_{nm}/(1+\rho_{kl})$, we get the second inequality. If $kl = 0$ or $kl = nm$, we can prove this in a similar way. Lemma 5.1 (5.7) follows.

In order to get the third inequality, for $0 < kl < nm$ and for $h \geq 0$, we have

$$\begin{aligned}
&E_0\big(\exp\big(-b^2 + b(Y_{C_1} + Y_{C_2})\big)\mathbb{1}_{Y_V \leq T_{kl}}\big) \\
&\quad = E_0\big(\exp\big(-c^2/2 + cY_{V_1}\big)\big) \cdot E_0\big(\exp\big(-c^2/2 + cY_{V_2}\big)\big) \cdot E_0\big(\mathrm{e}^{-z_{kl}^2} \exp(2z_{kl}Y_V)\mathbb{1}_{Y_V \leq T_{kl}}\big) \\
&\quad \leq \mathrm{e}^{-z_{kl}^2 + T_{kl}h} E_0\big(\exp\big((2z_{kl} - h)Y_V + h(Y_V - T_{kl})\big)\mathbb{1}_{Y_V \leq T_{kl}}\big) \\
&\quad \leq \mathrm{e}^{-z_{kl}^2 + T_{kl}h} E_0\big(\exp\big((2z_{kl} - h)Y_V\big)\big) = \exp\big(-z_{kl}^2 + T_{kl}h + (2z_{kl} - h)^2/2\big).
\end{aligned}$$

Taking $h = 2z_{kl} - T_{kl}$, we get the third inequality. If $kl = nm$, we can prove this in a similar way. Lemma 5.1 (5.8) follows.

A.6. Proof of Lemma 5.3

Recalling $P_{n,p}(k) = \binom{n}{k} p^k (1-p)^{n-k}$. Using well-known inequality $\binom{n}{k} \leq (ne/k)^k$, we get

$$\log\big(P_{n,p}(k)\big) \leq k\big(\log(p) + \log(n/k) + 1\big).$$

Since $n/k \leq r(p)$, we see that $0 \leq \log(n/k) \leq \log(r(p)) = \mathrm{o}(\log(p^{-1}))$ under the assumption on $r(p)$. This implies the first relation of Lemma 5.3.

In view of Lemma 5.2, we have

$$P_{N,n,n}(k) \leq P_{\mathcal{HG}}(Z \geq k) \leq P_{\mathrm{Bin}}(Z \geq k) \sim P_{n,p}(k).$$

Lemma 5.3 follows.

A.7. Proof of Theorem 4.1

Let us see that $E_0(\hat{\sigma}^2) = \sigma^2$ and that $\mathrm{Var}_0(\hat{\sigma}^2) = 2\sigma^4/(NM)$. Denote by

$$O_B = \left\{ \left| \frac{\hat{\sigma}^2}{\sigma^2} - 1 \right| \leq \frac{B}{\sqrt{NM}} \right\},$$

with $B \to \infty$ such that $B/\sqrt{NM} \to 0$. Then, $P_0((O_B)^c) \leq 2B^{-2} = \mathrm{o}(1)$.

It is easy to see that $P_0(\hat{t}_{\mathrm{lin}} > H) \leq P_0(t_{\mathrm{lin}}/\sigma > H\hat{\sigma}/\sigma, O_B) + P_0((O_B)^c) = P_0(t_{\mathrm{lin}}/\sigma > \tilde{H}) + \mathrm{o}(1) = \mathrm{o}(1)$, as $\tilde{H}^2 = H^2(1 - B/\sqrt{NM}) \to \infty$.

Similarly, for $\hat{t}_{\max}$, we put $T_{nm,\delta} = T_{nm}\sqrt{1 + \delta/2} =$ and $\tilde{T}_{nm}^2 = T_{nm,\delta}^2(1 - B\sqrt{NM})$ and then $P_0(\hat{t}_{\max} > T_{nm,\delta}) \leq P_0(t_{\max}/\sigma > \tilde{T}_{nm}) + P_0((O_B)^c) = \mathrm{o}(1)$, for our choice of $T_{nm,\delta}$. These imply $\alpha(\hat{\psi}^*) \to 0$.

Under the alternative, let us see that

$$\beta\big(\hat{\psi}^*, S\big) \leq \min\big\{P_S(\hat{t}_{\mathrm{lin}} \leq H), P_S(\hat{t}_{\max} \leq T_{nm,\delta})\big\},$$

and it suffices to check that either $P_S(\hat{t}_{\text{lin}} \le H) \to 0$ or $P_S(\hat{t}_{\max} \le T_{nm,\delta}) \to 0$. We can decompose

$$\hat{\sigma}^2 = \frac{\sigma^2}{NM}\biggl(\sum_{(i,j)\notin C} \xi_{ij}^2 + \sum_{(i,j)\in C} (s_{ij} + \xi_{ij})^2\biggr)$$
$$= \frac{\sigma^2}{NM}\biggl(\sum_{(i,j)} \xi_{ij}^2 + 2\sum_{(i,j)\in C} \frac{s_{ij}}{\sigma}\xi_{ij} + \sum_{(i,j)\in C} \frac{s_{ij}^2}{\sigma^2}\biggr).$$

We get

$$E_S(\hat{\sigma}^2) = \sigma^2(1+G), \qquad \mathrm{Var}_S(\hat{\sigma}^2) = \frac{2\sigma^4}{NM}(1+2G) \qquad \text{where } G = \frac{1}{\sigma^2 NM}\sum_{(i,j)\in C} s_{ij}^2.$$

Denote by $R := B\sqrt{\frac{2}{NM}(1+2G)}$ with $B \to \infty$ such that $B = \mathrm{o}(\sqrt{NM})$, and by

$$O_{SB} = \biggl\{\biggl|\frac{\hat{\sigma}^2}{\sigma^2} - 1 - G\biggr| \le R\biggr\}.$$

Then $P_S((O_{SB})^c) \le B^{-2} = \mathrm{o}(1)$ and $R = \mathrm{o}(1+G)$. Recalling

$$E_S(t_{\text{lin}}/\sigma) = (NM)^{-1/2}\sum_{(i,j)\in C} s_{ij}/\sigma \ge a\sqrt{nmpq}/\sigma,$$

we see that

$$G \le \frac{\max_{(i,j)\in C} s_{ij}}{\sum_{(i,j)\in C} s_{ij}} \frac{(\sum_{(i,j)\in C} s_{ij})^2}{\sigma^2 NM} = \tau_G\bigl(E_S(t_{\text{lin}}/\sigma)\bigr)^2.$$

This implies

$$1 + G + R \le \bigl(1 + \tau_G\bigl(E_S(t_{\text{lin}}/\sigma)\bigr)^2\bigr)\bigl(1+\mathrm{o}(1)\bigr). \tag{A.5}$$

Let $E_S(t_{\text{lin}}/\sigma) \to \infty$, which holds when $a\sqrt{nmpq}/\sigma \to \infty$. By our choice of H, (4.1) and (A.5) we have $H\sqrt{1+G+R} = \mathrm{o}(E_S(t_{\text{lin}}/\sigma))$. Applying the Chebyshev inequality and since $\mathrm{Var}_S(t_{\text{lin}}/\sigma) = 1$, we have

$$\begin{aligned}
P_S(\hat{t}_{\text{lin}} \le H) &\le P_S(t_{\text{lin}}/\sigma \le H\hat{\sigma}/\sigma, O_{SB}) + P_S\bigl((O_{SB})^c\bigr)\\
&\le P_S(t_{\text{lin}}/\sigma \le H\sqrt{1+G+R}) + \mathrm{o}(1)\\
&\le P_S\bigl(E_S(t_{\text{lin}}/\sigma) - t_{\text{lin}}/\sigma \le E_S(t_{\text{lin}}/\sigma) - H\sqrt{1+G+R}\bigr) + \mathrm{o}(1)\\
&\le \bigl(E_S(t_{\text{lin}}/\sigma) - H\sqrt{1+G+R}\bigr)^{-2} + \mathrm{o}(1) = \mathrm{o}(1).
\end{aligned}$$

This proves that $P_S(\hat{t}_{\text{lin}} \le H) \to 0$ if $E_S(t_{\text{lin}}/\sigma) \to \infty$.

If $E_S(t_{\text{lin}}/\sigma) = \mathrm{O}(1)$ (this is possible when $a\sqrt{nmpq}/\sigma = \mathrm{O}(1)$ only), we have $\sqrt{1+G+R} = 1 + \mathrm{o}(1)$ by (4.1) and (A.5). Therefore

$$P_S(\hat{t}_{\max} \le T_{nm,\delta}) \le P_S(t_{\max} \le T_{nm,\delta}\hat{\sigma}/\sigma, O_{SB}) + P_S\big((O_{SB})^c\big)$$
$$\le P_S(t_{\max} \le T_{nm,\delta}\sqrt{1+G+R}) + \mathrm{o}(1) = \mathrm{o}(1)$$

for our choice of $T_{nm,\delta}$ and by the second assumption (4.2) (compare with the proof of Theorem 2.1). These implies $\beta_{nm,a}(\hat{\psi}^*) \to 0$. Thus $\gamma_{nm,a}(\hat{\psi}^*) \to 0$.

A.8. Proof of the upper bound of Theorem 4.2

It follows the same lines as that of Theorem 2.1. We use Markov inequality and bound from above exponential moments of our test statistics (as they are not having Gaussian distribution in this case).

We use repeatedly the well-known facts that, $A'(s^0) = 0$ and $A''(s^0) = 1$ for centered and reduced random variable at s^0. Moreover,

$$E_s\big[\mathrm{e}^{uY}\big] = \mathrm{e}^{A(s+u)-A(s)}$$

for any s and u such that s ans $s+u$ are interior points of the parameter space. For the statistic t_{lin}, we have

$$\alpha\big(\psi_H^{\text{lin}}\big) = P_{s^0}\big(t_{\text{lin}}H > H^2\big) \le \mathrm{e}^{-H^2} E_{s^0}\big[\mathrm{e}^{t_{\text{lin}}H}\big]$$
$$\le \mathrm{e}^{-H^2} \prod_{(i,j)} E_{s^0}\big[\mathrm{e}^{Y_{ij}H/\sqrt{NM}}\big]$$
$$\le \exp\left(\left(A\left(s^0 + \frac{H}{\sqrt{NM}}\right) - A\big(s^0\big)\right)NM - H^2\right).$$

For $H \to \infty$ as in our theorem $H/\sqrt{NM} \to 0$, then we get $A(s^0 + \frac{H}{\sqrt{NM}}) - A(s^0) = \frac{H^2}{2NM}(1 + \mathrm{o}(1))$ and $\alpha(\psi_H^{\text{lin}}) \le c\mathrm{e}^{-H^2/2(1-\mathrm{o}(1))} \to 0$, for some constant $c > 0$.

Under the alternative,

$$\beta\big(\psi_H^{\text{lin}}, S\big) = P_S[-t_{\text{lin}} + H \ge 0] \le \mathrm{e}^H E_S\big[\mathrm{e}^{-t_{\text{lin}}}\big]$$
$$\le \mathrm{e}^H \prod_{(i,j)\notin C} E_{s^0}\big[\mathrm{e}^{-Y_{ij}/\sqrt{NM}}\big] \prod_{(i,j)\in C} E_{s_{ij}}\big[\mathrm{e}^{-Y_{ij}/\sqrt{NM}}\big]$$
$$\le \mathrm{e}^{H+(NM-nm)(A(s^0-(1/\sqrt{NM}))-A(s^0))} \prod_{(i,j)\in C} \mathrm{e}^{A(s_{ij}-(1/\sqrt{NM}))-A(s_{ij})}.$$

On the one hand, $A(s^0 - \frac{1}{\sqrt{NM}}) - A(s^0) \sim -\frac{1}{2NM}$. On the other hand, it is easy to check that $A''(s) \ge 0$. Thus, A is a convex function and A' is increasing. This implies $A(s_{ij} - \frac{1}{\sqrt{NM}}) - A(s_{ij}) \le -\frac{1}{\sqrt{NM}}A'(s_{ij})$ which is less or equal than $-\frac{1}{\sqrt{NM}}A'(s^0 + a)$ under the alternative.

Finally,

$$\beta\big(\psi_H^{\mathrm{lin}}, S\big) \le \exp\biggl(H - \frac{1}{2}(1-pq) - A'\big(s^0+a\big)\frac{nm}{\sqrt{NM}}\biggr)$$
$$\le c\exp\big(H - A'\big(s^0+a\big)\sqrt{nmpq}\big) \to 0$$

under our assumption.

For the statistic $t_{\max}$,

$$\alpha\big(\psi^{\max}\big) = P_{s^0}(t_{\max} > T_{nm,\delta}) \le \sum_C P_{s^0}\big(t_{\max} T_{nm,\delta} > T_{nm,\delta}^2\big)$$
$$\le \sum_C \mathrm{e}^{-T_{nm,\delta}^2} E_{s^0}\big[\mathrm{e}^{Y_C T_{nm,\delta}}\big] \le \sum_C \mathrm{e}^{-T_{nm,\delta}^2} E_{s^0}\big[\mathrm{e}^{\sum_{(i,j)\in C} Y_{ij} T_{nm,\delta}/\sqrt{nm}}\big]$$
$$\le G_{nm}\mathrm{e}^{-T_{nm,\delta}^2}\mathrm{e}^{(A(s^0+T_{nm,\delta}/\sqrt{nm})-A(s^0))\times nm}.$$

As $T_{nm,\delta}/\sqrt{nm} \asymp (\log(p^{-1})/m + \log(q^{-1})/n)^{1/2} \to 0$, we have $A(s^0+T_{nm,\delta}/\sqrt{nm}) - A(s^0) \asymp T_{nm,\delta}^2/(2nm)$ and this gives

$$\alpha\big(\psi^{\max}\big) \le G_{nm}\mathrm{e}^{-(T_{nm,\delta}^2/2)(1-\mathrm{o}(1))} \to 0.$$

For the second-type error,

$$\beta\big(\psi^{\max}, S\big) \le P_S[-Y_C + T_{nm,\delta} \ge 0] \le \mathrm{e}^{T_{nm,\delta}} E_S\big[\mathrm{e}^{-Y_C}\big]$$
$$\le \exp(T_{nm,\delta}) \cdot \prod_{(i,j)\in C} E_{s_{ij}}\big[\mathrm{e}^{-Y_{ij}/\sqrt{nm}}\big]$$
$$\le \exp\biggl(T_{nm,\delta} + \sum_{(i,j)\in C}\biggl[A\biggl(s_{ij} - \frac{1}{\sqrt{nm}}\biggr) - A(s_{ij})\biggr]\biggr)$$
$$\le \exp\biggl(T_{nm,\delta} - nmA'\big(s^0+a\big)\frac{1}{\sqrt{nm}}\biggr) \to 0,$$

by the choice of $\delta > 0$ small enough.

A.9. Proof of Theorem 4.3

A.9.1. *Proof of the upper bounds*

We have, under the null hypothesis, $z_{\mathrm{lin}}\sqrt{2NM}$ has a χ^2 distribution with $E_0(z_{\mathrm{lin}}) = 0$ and $\mathrm{Var}_0(z_{\mathrm{lin}}) = 1$. This implies that $P_0(z_{\mathrm{lin}} > H) \to 0$ as $H \to \infty$.

For $z_{\max}$ we will use the moment generating function of the χ^2 distribution. We have

$$P_0(z_{\max} > T_{nm,\delta}) \le \sum_C P_0(Z_C > T_{nm,\delta}) \le G_{nm} P_0\big(T_{nm,\delta} Z_C > T_{nm,\delta}^2\big)$$
$$\le G_{nm}\mathrm{e}^{-T_{nm,\delta}^2} E_0\big(\mathrm{e}^{T_{nm,\delta} Z_C}\big)$$

$$\leq G_{nm} \mathrm{e}^{-T_{nm,\delta}^2 - T_{nm,\delta}\sqrt{nm/2}} \left(1 - \frac{2T_{nm,\delta}}{\sqrt{2nm}}\right)^{-nm/2}$$

$$\leq G_{nm} \mathrm{e}^{-(T_{nm,\delta}^2/2)(1+\mathrm{o}(1))} = \mathrm{o}(1),$$

by the choice of $T_{nm,\delta}$ in our theorem. Indeed, $2T_{nm,\delta}^2/(nm) = \mathrm{O}(a^4) \to 0$, by assumption (4.6).

Again, $\beta(\psi^z, S) \leq \min\{P_S(z_{\mathrm{lin}} \leq H), P_S(z_{\max} \leq T_{nm,\delta})\}$. Under the alternative, $S = S_C$ and z_{lin} has mean $E_S(z_{\mathrm{lin}}) = \lambda/\sqrt{NM}$ and variance $\mathrm{Var}_S(z_{\mathrm{lin}}) = 1 + 2\lambda/(NM)$, where $\lambda = \sum_C s_{ij}^2$. We have,

$$\frac{\lambda}{\sqrt{2NM}} \geq \frac{a^2 nm}{\sqrt{2NM}} \geq \frac{a^2}{\sqrt{2}} \sqrt{nmpq}.$$

Therefore, if $a^2\sqrt{nmpq} \to \infty$, we have

$$P_S(z_{\mathrm{lin}} \leq H) \leq \frac{\mathrm{Var}_S(z_{\mathrm{lin}})}{(E_S(z_{\mathrm{lin}}) - H)^2} = \frac{1 + 2\lambda/(NM)}{(1-c)^2\lambda^2/(2NM)}$$

$$\leq \frac{2}{(1-c)^2 a^4 nmpq} + \frac{4}{(1-c)^2 a^2 nm} = \mathrm{o}(1).$$

Under the alternative,

$$P_S(z_{\max} \leq T_{nm,\delta}) \leq \mathrm{e}^{T_{nm,\delta}} E_{S_C}\left[\mathrm{e}^{-Z_C}\right] \leq \mathrm{e}^{T_{nm,\delta} + \sqrt{nm/2}} E_{S_C}\left[\exp\left(t \sum_C Y_{ij}^2\right)\right],$$

where $t = -1/\sqrt{2nm} < 1/2$. Therefore,

$$E_{S_C}\left[\exp\left(t \sum_C Y_{ij}^2\right)\right] = \exp\left(\frac{\lambda t}{1-2t} - \frac{nm}{2} \log(1-2t)\right)$$

$$\leq \exp\left(-\frac{\lambda}{\sqrt{2nm}} \frac{1}{1+\sqrt{2/(nm)}} - \frac{nm}{2}\left(\sqrt{\frac{2}{nm}} - \frac{1}{nm}\right)\right)$$

$$\leq \exp\left(-\frac{a^2\sqrt{nm}}{\sqrt{2}} \frac{1}{1+\sqrt{2/(nm)}} - \sqrt{\frac{nm}{2}} + \frac{1}{2}\right).$$

In conclusion, if $\liminf a^4 nm/(4(n\log(p^{-1}) + m\log(q^{-1}))) > 1$ we have

$$P_{S_C}(z_{\max} \leq T_{nm,\delta}) \leq \sqrt{e} \exp\left(T_{nm,\delta} - \frac{a^2\sqrt{nm}}{\sqrt{2}} \frac{1}{1+\sqrt{2/(nm)}}\right) \to 0.$$

A.9.2. *Proof of the lower bounds*

We follow the lines of the proof of Theorem 2.2. The prior on the set of matrices is $\pi = G_{nm}^{-1} \sum_{C \in \mathcal{C}_{nm}} \pi_C$, where, under π_C, the matrix $S = S_C$ has $s_{ij} = 0$ with probability 1 for all $(i, j) \notin C$ and s_{ij} is either a and $-a$ with probability $1/2$, for all $(i, j) \in C$.

Let P_{S_C} denote the likelihood of the random variables in $\mathbf{Y}$ when $S = S_C$ and P_π denote the mixture of likelihoods $P_\pi = G_{nm}^{-1} \sum_{C\in\mathcal{C}_{nm}} P_{S_C}$. Therefore, the likelihood ratio $L_\pi(Y)$ is

$$\begin{aligned} L_\pi(Y) &= \frac{\mathrm{d}P_\pi}{\mathrm{d}P_0}(Y) = \frac{1}{G_{nm}} \sum_{C\in\mathcal{C}_{nm}} \frac{\mathrm{d}P_{\pi_C}}{\mathrm{d}P_0}(Y) \\ &= \frac{1}{G_{nm}} \sum_{C\in\mathcal{C}_{nm}} \prod_{(i,j)\in C} \mathrm{e}^{-a^2/2} \cosh(aY_{ij}) \\ &= \frac{1}{G_{nm}} \sum_{C\in\mathcal{C}_{nm}} \mathrm{e}^{-a^2 nm/2} \exp\Biggl(\sum_{(i,j)\in C} \log\bigl(\cosh(aY_{ij})\bigr)\Biggr). \end{aligned}$$

Note that $E_0[\cosh(aY_{ij})] = \mathrm{e}^{a^2/2}$ and that $E_0[\cosh^2(aY_{ij})] = (1 + \mathrm{e}^{2a^2})/2$. We can reproduce the proof of Theorem 2.2 with a^2 replaced by $a^4/2$. For example, in (5.6) we have

$$g(k,l) = E_0\biggl[\exp\biggl(-a^2 nm + \sum_{C_1} \log\bigl(\cosh(aY_{ij})\bigr) + \sum_{C_2} \log\bigl(\cosh(aY_{ij})\bigr)\biggr) \mathbb{1}_{\Gamma_{C_1}\cap\Gamma_{C_2}}\biggr].$$

We can show, as in the proof of (5.6), that

$$\begin{aligned} g(k,l) &\le \mathrm{e}^{-a^2 nm} E_0\biggl[\exp\biggl(\sum_{V_1\cup V_2} \log\bigl(\cosh(aY_{ij})\bigr) + 2\sum_{V} \log\bigl(\cosh(aY_{ij})\bigr)\biggr)\biggr] \\ &\le \mathrm{e}^{-a^2 nm} \mathrm{e}^{2(a^2/2)(nm-kl)} E_0^{kl}\bigl[\cosh^2(aY)\bigr] \le \mathrm{e}^{-a^2 kl} \biggl(\frac{1+\mathrm{e}^{2a^2}}{2}\biggr)^{kl} \\ &= \bigl(\cosh\bigl(a^2\bigr)\bigr)^{kl} \le \mathrm{e}^{(a^4/2)kl}, \end{aligned}$$

where V_1, V_2 and V are defined in the proof of Lemma 5.1.

The relations (5.7) and (5.8) could be replaced by the following:

$$g(k,l) \le \exp\biggl(-(T_{nm} - b)^2 + \frac{\rho_{kl} T_{nm}^2}{1+\rho_{kl}} + \mathrm{o}\bigl(T_{nm}^2\bigr)\biggr), \tag{A.6}$$

$$g(k,l) \le \exp\bigl(T_{kl}^2/2 - (T_{kl} - z_{kl})^2 + \mathrm{o}\bigl(T_{kl}^2\bigr)\bigr), \tag{A.7}$$

where $b^2 = nma^4/2$, $z_{kl}^2 = b^2\rho_{kl}$, under the same constraints. The inspection of the proofs of (5.7) and (5.8) shows that, in order to prove (A.6) and (A.7), one could use the following relation in place of (A.4):

$$E_0\bigl[\mathrm{e}^{\tau(\log(\cosh(aY)) - a^2/2 + a^4/4)}\bigr] = \exp\biggl(\frac{\tau^2 a^4}{4} + \mathrm{o}\bigl(a^4\bigr)\biggr) \tag{A.8}$$

for $a \to 0$ and $\tau \in \mathbb{R}^+$, $\tau = \mathrm{O}(1)$.

In order to prove (A.8), we can split the expected value over the events $\{\tau a^2 Y^2 > \delta^2\}$ and $\{\tau a^2 Y^2 \le \delta^2\}$, respectively, for some small enough $\delta > 0$ such that $\delta/a\sqrt{\tau} \to \infty$ (we choose $\delta = (\tau a^2)^{1/4}$). Firstly, we use the inequality $\cosh(x) \le \mathrm{e}^{x^2/2}$ and get

$$\begin{aligned}
E_0\bigl[\mathrm{e}^{\tau \log(\cosh(aY))} \cdot \mathbb{1}_{\tau a^2 Y^2 > \delta^2}\bigr] &\le E_0\bigl[\mathrm{e}^{\tau a^2 Y^2/2} \cdot \mathbb{1}_{\tau a^2 Y^2 > \delta^2}\bigr] \\
&\le 2 \int_{\delta/a\sqrt{\tau}}^{\infty} \mathrm{e}^{-(1-\tau a^2) y^2/2} \frac{\mathrm{d}y}{\sqrt{2\pi}} \\
&\le \sqrt{\frac{2}{\pi}} \frac{\delta/(a\sqrt{\tau})}{1-\tau a^2} \exp\biggl(-\frac{1-\tau a^2}{2} \frac{\delta^2}{\tau a^2}\biggr) = \mathrm{o}\bigl(\tau^2 a^4\bigr).
\end{aligned}$$

Secondly, on the event $\{\tau a^2 Y^2 \le \delta^2\}$ we use the Taylor expansions $\log(\cosh(x)) = x^2/2 - x^4/12(1 + \mathrm{o}(1))$, $\mathrm{e}^x = 1 + x + x^2/2(1 + \mathrm{o}(1))$, $x = \mathrm{o}(1)$. Denote $U = \log(\cosh(aY)) - E_0(\log(\cosh(aY)))$. We have

$$E_0\bigl(\log\bigl(\cosh(aY)\bigr)\bigr) = \frac{a^2}{2} - \frac{a^4}{4}\bigl(1 + \mathrm{o}(1)\bigr), \qquad \mathrm{Var}_0(U) = \frac{a^4}{2}\bigl(1 + \mathrm{o}(1)\bigr), \tag{A.9}$$

and, since $\tau U = \mathrm{o}(1)$,

$$\begin{aligned}
E_0\bigl[\mathrm{e}^{\tau U} \cdot \mathbb{1}_{\tau a^2 Y^2 \le \delta^2}\bigr] &= E_0\biggl[\biggl(1 + \tau U + \frac{\tau^2 U^2}{2}\bigl(1 + \mathrm{o}(1)\bigr)\biggr) \cdot \mathbb{1}_{\tau a^2 Y^2 \le \delta^2}\biggr] \\
&= 1 + \frac{\tau^2 \mathrm{Var}_0(U)}{2}\bigl(1 + \mathrm{o}(1)\bigr) \\
&\quad - E_0\biggl[\biggl(1 + \tau U + \frac{\tau^2 U^2}{2}\bigl(1 + \mathrm{o}(1)\bigr)\biggr) \cdot \mathbb{1}_{\tau a^2 Y^2 > \delta^2}\biggr].
\end{aligned}$$

The last expected value is $\mathrm{o}(\tau^2 a^4)$ and this gives

$$E_0\bigl[\mathrm{e}^{\tau \log(\cosh(aY))} \cdot \mathbb{1}_{\tau a^2 Y^2 \le \delta^2}\bigr] = \mathrm{e}^{\tau E_0[\log\cosh(aY)] + (\tau^2 a^4/4)(1+\mathrm{o}(1))}.$$

Together with the first relation in (A.9), this ends the proof of (A.8).

A.10. Proof of Theorem 4.4

A.10.1. *Proof of the lower bounds*

Let $K = [N/n]$, $L = [M/m]$ and consider only non-overlapping rectangles

$$R_{kl} = \bigl\{(i,j)\colon\ n(k-1) + 1 \le i \le nk, m(l-1) + 1 \le j \le ml\bigr\}, \qquad 1 \le k \le K, 1 \le l \le L.$$

Let S_{kl} be the matrix with the elements $s_{ij} = 0$ if $(i, j) \notin R_{kl}$ and $s_{ij} = a$ if $(i, j) \in R_{kl}$. Consider the prior

$$\pi = \frac{1}{KL} \sum_{k=1}^{K} \sum_{l=1}^{L} \delta_{S_{kl}}.$$

By construction, $\pi(\{S_{kl}, k, l\}) = 1$. The likelihood ratio is of the form

$$L(Y) = \frac{\mathrm{d}P_\pi}{\mathrm{d}P_0}(Y) = \frac{1}{KL} \sum_{k=1}^{K} \sum_{l=1}^{L} \frac{\mathrm{d}P_{S_{kl}}}{\mathrm{d}P_0}(Y) = \frac{1}{KL} \sum_{k=1}^{K} \sum_{l=1}^{L} \exp\bigl(-b^2/2 + bZ_{kl}\bigr),$$

where

$$Z_{kl} = \frac{1}{\sqrt{nm}} \sum_{(i,j)\in R_{kl}} Y_{ij}, \qquad b^2 = nma^2.$$

Note that $Z_{kl} \sim \mathcal{N}(0, 1)$ under P_0 and are independent in k, l. It is sufficient to check that $L(Y) \to 1$ in P_0-probability. Let us consider the truncated likelihood ratio

$$\tilde{L}(Y) = \frac{1}{KL} \sum_{k=1}^{K} \sum_{l=1}^{L} \exp\bigl(-b^2/2 + bZ_{kl}\bigr) \mathbb{1}_{Z_{kl} < T_{KL}},$$

where we set

$$T_{KL} = \sqrt{2\log(KL)} \sim \sqrt{2\bigl(\log\bigl(p^{-1}\bigr) + \log\bigl(q^{-1}\bigr)\bigr)}.$$

Since

$$P_0(L \neq \tilde{L}) \le \sum_{k=1}^{K} \sum_{l=1}^{L} P_0(Z_{kl} \ge T_{KL}) \to 0,$$

it suffices to check that $\tilde{L}(Y) \to 1$ in P_0-probability.

Observe now that $T_{KL} - b \to \infty$ under the assumptions of the theorem, and it suffices to consider the case $b > cT_{kl}$ for some $c \in (1/2, 1)$. We have

$$E_0\bigl(\tilde{L}(Y)\bigr) = \frac{1}{KL} \sum_{k=1}^{K} \sum_{l=1}^{L} E_0\bigl(\exp\bigl(-b^2/2 + bZ_{kl}\bigr)\mathbb{1}_{Z_{kl}<T_{KL}}\bigr) = \Phi(T_{KL} - b) \to 1,$$

$$\begin{aligned}
\mathrm{Var}_0\bigl(\tilde{L}(Y)\bigr) &= \frac{1}{(KL)^2} \sum_{k=1}^{K} \sum_{l=1}^{L} \mathrm{Var}_0\bigl(\exp\bigl(-b^2/2 + bZ_{kl}\bigr)\mathbb{1}_{Z_{kl}<T_{KL}}\bigr) \\
&\le \frac{1}{(KL)^2} \sum_{k=1}^{K} \sum_{l=1}^{L} E_0\bigl(\exp\bigl(-b^2 + 2bZ_{kl}\bigr)\mathbb{1}_{Z_{kl}<T_{KL}}\bigr) \\
&= \frac{1}{KL} \exp\bigl(b^2\bigr)\Phi(T_{KL} - 2b)
\end{aligned}$$

$$\leq \exp\big(b^2 - (T_{KL} - 2b)^2/2 - T_{KL}^2/2\big)$$
$$= \exp\big(-(T_{KL} - b)^2\big) \to 0.$$

Proof of the lower bounds in Theorem 4.4.

A.10.2. *Proof of the upper bounds*

Set $T_{KL} = \sqrt{2\log(KL)}$ and observe that, by the choice of η and since $pq \to 0$, we have

$$\begin{aligned} T_{KL} &= \sqrt{2\log(NM/nm) - 4\log(\eta) + \mathrm{o}(1)} \\ &= \big(2\log\big((pq)^{-1}\big)\big)^{1/2}\big(1 + \big(\log(\eta) + \mathrm{o}(1)\big)/\log\big((pq)^{-1}\big)\big)^{1/2} \\ &\sim \sqrt{2\big(\log\big(p^{-1}\big) + \log\big(q^{-1}\big)\big)}. \end{aligned}$$

For type I errors, we have

$$\alpha(\psi_Z) \leq \sum_{k=1}^{K}\sum_{l=1}^{L} P_0(Z_{kl} > T_{KL}) = KL\Phi(-T_{KL}) \to 0.$$

Let the alternative S_E correspond to the matrix with entry $a > 0$ at positions in $E = E_{k^*l^*}$ and 0 elsewhere. As previously, $E = E_{k^*l^*}, 0 \leq k^* \leq N - n, 0 \leq l^* \leq M - m$ consists of (i, j) such that $k^* < i \leq k^* + n, l^* < i \leq l^* + m$. By construction, we can take $k, l, 1 \leq k \leq K, 1 \leq l \leq L$ such that $|n_k - k^*| \leq n\eta$, $|m_l - l^*| \leq m\eta$. Therefore, the matrix $E_{k^*l^*}$ will overlap with the matrix $E_{n_k m_l}$ from our test procedure significantly:

$$\begin{aligned} \tilde{n} &= \#\big(\big\{k^* + 1, \ldots, k^* + n\big\} \cap \{n_k + 1, \ldots, n_k + n\}\big) \geq n(1 - \eta), \\ \tilde{m} &= \#\big(\big\{l^* + 1, \ldots, l^* + m\big\} \cap \{m_l + 1, \ldots, m_l + m\}\big) \geq m(1 - \eta). \end{aligned}$$

Observe that

$$\beta(\psi_Z, S_E) \leq P_{S_E}(Z_{kl} \leq T_{KL}).$$

Moreover, $Z_{kl} \sim \mathcal{N}(\tilde{b}, 1)$ under P_{S_E} where we recall that $b = a\sqrt{nm}$ and we put

$$\tilde{b} = \frac{a\tilde{n}\tilde{m}}{\sqrt{nm}} \geq b(1 - \eta)^2 \sim b.$$

This yields

$$\beta(\psi_Z, S_E) \leq \Phi\big(T_{KL} - b\big(1 + \mathrm{o}(1)\big)\big) \to 0$$

under assumptions of theorem. Proof of the upper bounds in Theorem 4.4 follows.

Acknowledgements

Research was partially supported by RFBR Grant 11-01-00577 and by the Grant NSh–4472.2010.1. The author acknowledges support from the CNRS for his visit to the University Paris-Est Marne-la-Vallée.

References

[1] Addario-Berry, L., Broutin, N., Devroye, L. and Lugosi, G. (2010). On combinatorial testing problems. *Ann*. *Statist*. **38** 3063–3092. MR2722464

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

[3] Arias-Castro, E., Candès, E.J. and Plan, Y. (2010). Global testing and sparse alternatives: ANOVA, multiple comparisons and the higher criticism. Available at arXiv:1007.1434.

[4] Arias-Castro, E., Donoho, D.L. and Huo, X. (2005). Near-optimal detection of geometric objects by fast multiscale methods. *IEEE Trans*. *Inform*. *Theory* **51** 2402–2425. MR2246369

[5] Bickel, P.J., Ritov, Y. and Tsybakov, A.B. (2009). Simultaneous analysis of lasso and Dantzig selector. *Ann*. *Statist*. **37** 1705–1732. MR2533469

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

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

[8] Ingster, Y.I. and Suslina, I.A. (2002). On a detection of a signal of known shape in multichannel system. *Zapiski Nauchn*. *Sem*. *POMI* **294** 88–112 (in Russian) (transl. *J. Math. Sci.* **127** 1723–1736).

[9] Koltchinskii, V., Lounici, K. and Tsybakov, A.B. (2011). Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion. *Ann*. *Statist*. **39** 2302–2329. MR2906869

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

[11] Sun, X. and Nobel, A.B. (2010). On the maximal size of large-average and ANOVA-fit submatrices in a Gaussian random matrix. Available at arXiv:1009.0562v1.

Received September 2011 and revised July 2012