Detecting Positive Correlations in a Multivariate Sample

Ery Arias-Castro* Sébastien Bubeck† Gábor Lugosi‡

February 24, 2012

Abstract

We consider the problem of testing whether a correlation matrix of a multivariate normal population is the identity matrix. We focus on sparse classes of alternatives where only a few entries are nonzero and, in fact, positive. We derive a general lower bound applicable to various classes and study the performance of some near-optimal tests. We pay special attention to computational feasibility and construct near-optimal tests that can be computed efficiently. Finally, we apply our results to prove new lower bounds for the clique number of high-dimensional random geometric graphs.

Keywords: sparse covariance matrices, sparse detection, high-dimensional data, minimax detection, Bayesian detection, random geometric graphs.

1 Introduction

In multivariate statistics, inference about a covariance (i.e., dispersion) matrix aims at answering questions of dependencies between the variables. This is strictly true when the variables are jointly Gaussian, which is the classical assumption. A basic question is whether the variables are dependent at all. Concretely, consider a simple setting where the components of a random vector are jointly normal, each with zero mean and unit variance. Then the variables are independent if and only if their covariance matrix is the identity matrix. As usual, inference is based on an i.i.d. sample of size m, denoted $X_1, \dots, X_m$ with $X_t = (X_{t,1}, \dots, X_{t,n}) \in \mathbb{R}^n$ for $t = 1, \dots, m$. As stated above, we assume that $\mathbb{E} X_{t,i} = 0$ and $\mathrm{Var}(X_{t,i}) = 1$, and let $\sigma_{i,j} = \mathrm{Cov}(X_{t,i}, X_{t,j})$.

We are interested in testing whether the population covariance matrix is the identity matrix, or not, so the null hypothesis is

$$H_0 : \sigma_{i,j} = 0, \ \forall i \neq j \ .$$

This testing problem is well studied in the classical regime where the dimension n is fixed and the sample size m increases to infinity, see Muirhead (1982, Sec. 8.4). Here, we study the regime where the dimension is large, that is, $n \to \infty$, and focus on alternatives where

*Department of Mathematics, University of California, San Diego
†Department of Operations Research and Financial Engineering, Princeton University
‡ICREA and Department of Economics, Universitat Pompeu Fabra

the covariance matrix is sparse in the usual sense, meaning that even under the alternative, only a few variables are substantially correlated.

This sparse setting has been investigated in the last few years, with recent work on the estimation of sparse covariance matrices Bickel and Levina (2008a,b); Cai et al. (2010); El Karoui (2008) and inference on sparse graphical models, for example Lam and Fan (2009); Meinshausen and Bühlmann (2006); Rajaratnam et al. (2008); Verzelen and Villers (2010); Yuan and Lin (2007). While the literature has focused almost exclusively on estimating the dependency structure, we focus here on the more basic task of testing whether there is any dependency at all. In this line of work, we find Chen et al. (2010); Ledoit and Wolf (2002), where modified versions of the likelihood ratio test are proposed to handle the case where the dimension n increases with the sample size. This is the setting we consider, specializing to the case of sparse correlation structures under the alternative.

1.1 Correlation models

We introduce sparse models of correlation matrices to test against. Though many more models are possible, we choose a few emblematic examples that are of interest in a much wider sense within the literature on sparse covariance estimation and on sparse graphical models. In all cases, the null hypothesis is that the observed vector has identity covariance matrix. For the alternative hypothesis, we consider the following prototypical examples:

- **Block model.** The covariance under the alternative is the identity matrix except for a $k \times k$ block on the diagonal. Formally, given $\rho > 0$, we assume here that there is a subset of indices of the form $S = \{i, \dots, i+k-1\}$ (called k-interval) such that $\sigma_{i,j} \geq \rho$ if $i, j \in S, i \neq j$. The set S is called the *anomalous set*.

- **Clique model.** This model is defined as the block model with the possible anomalous set S ranging over all the subsets of indices of size k (called k-set).

- **Perfect matching model.** Suppose n is a perfect square with $n = k^2$. Here the components of the observed vector X correspond to edges of the complete bipartite graph on $2k$ vertices. The alternative hypothesis is that the bipartite graph has a perfect matching such that $\sigma_{i,j} \geq \rho$ for all $i, j \in S, i \neq j$ where S is the anomalous set of indices corresponding to the edges of the perfect matching.

The block model is closely related to the models used in Cai et al. (2010) to obtain bounds on the minimax risk of estimating sparse matrices. Roughly speaking, Cai et al. (2010) use the block model with $S = \{1, \dots, k\}$ and place nonzero entries in a (carefully designed) fashion within that block. The fraction of nonzero entries within the block is about one-half. We could also assume that only a fraction of the entries in the block are nonzero and it would only change constants later on. More importantly, to make the detection problem interesting, we need to consider all possible blocks. Note that the block model is parametric. The clique model is a natural generalization of the block model leading to a nonparametric model. The perfect matching model gives an example of a class of sets with a more intricate combinatorial structure which our approach is able to deal with.

1.2 Tests and their risks

As usual, a *test* is a binary-valued function $f : \mathbb{R}^{nm} \to \{0, 1\}$, with $f(X_1, \ldots, X_m) = 1$ meaning that the test rejects the null H_0 in favor of the particular alternative of interest. We measure the performance of a test based on its *worst-case risk* over the model of interest $\mathcal{M}$, formally defined by

$$R^{\max}(f) = \mathbb{P}_0\{f(X_1, \ldots, X_m) = 1\} + \sup_{M \in \mathcal{M}} \mathbb{P}_M\{f(X_1, \ldots, X_m) = 0\} .$$

($\mathbb{P}_0$ denotes the distribution under the null while $\mathbb{P}_M$ denote the distribution under the alternative associated with a particular covariance structure M.) In our setup, $R^{\max}(f)$ depends on n, m, ρ, and the class $\mathcal{C}$ of possible index sets S. When all non-zero covariances $\sigma_{i,j}$ are actually equal to the lower bound ρ, then $\mathbb{P}_M$ is determined by S and, with a slight abuse of notation, we write $\mathbb{P}_S$ for $\mathbb{P}_M$. Clearly,

$$R^{\max}(f) \geq \mathbb{P}_0\{f(X_1, \ldots, X_m) = 1\} + \max_{S \in \mathcal{C}} \mathbb{P}_S\{f(X_1, \ldots, X_m) = 0\}$$

and indeed all lower bounds derived in this paper start with this inequality. We will derive upper and lower bounds for the *minimax risk*,

$$R_*^{\max} := \inf_f R^{\max}(f) ,$$

where the infimum is taken over all measurable functions $f : \mathbb{R}^{nm} \to \{0, 1\}$.

The lower bounds will be obtained by putting a prior on model $\mathcal{C}$ and obtaining a lower bound on the corresponding *Bayesian risk* which never exceeds the worst-case risk. In all cases, we draw the set S uniformly at random within the class $\mathcal{C}$. The upper bounds are obtained by studying the performance of specific tests.

We focus on the case where the dimension n and the sample size m are both large. Of course, such asymptotic statements only make sense if we define sequences of integers $m = m_n$, $k = k_n$, $\rho = \rho_n$ and classes $\mathcal{C} = \mathcal{C}_n$. This dependency in n will be left implicit. In this asymptotic setting, we say that *reliable* detection is possible (resp. impossible) if $R_*^{\max} \to 0$ (resp. $\to 1$) as $n \to \infty$. Also we say that a sequence of tests (f_n) is asymptotically powerful (resp. powerless) if $R^{\max}(f_n) \to 0$ (resp. $\to 1$).

1.3 A preview of results for the clique model

Among the models we consider, the clique model is perhaps the most compelling because of its relevance in applications and its complexity. Also, for a given value of k, the clique model is the richest possible and therefore for any given ρ, n, m, $R_*^{\max}$ is larger than for any other model. This makes the clique model an important benchmark.

Here we summarize our main findings for this special class. We discover various types of behavior in distinct ranges of the parameters m, k, ρ. Roughly speaking, and ignoring logarithmic factors, we arrive at the following conclusions. Two tests are competing for near-optimality. The first one is a 'global' test akin to the classical test (Muirhead, 1982, Sec. 8.4) and the refinements in Chen et al. (2010); Ledoit and Wolf (2002). The second is a 'local' test reminiscent to the generalized likelihood ratio test. The latter dominates the former when

$$\frac{\max\left(1,(k/m)^{1/2}\right)k^{3/2}}{n}$$

is small, meaning when the alternative is more sparse (corresponding to smaller values of k).

Our results also point out to an interesting, perhaps even surprising, phenomenon. It turns out that if the sample size m is $o(\log n)$, the risk of the optimal test is not significantly smaller than when $m=1$, that is, when only one observation is available. In other words, if the dimension n grows faster than exponential in the sample size m, the situation is essentially the same as if the sample size were equal to one. However, the situation becomes drastically different when m is much larger than $\log n$ as reliable detection becomes possible for significantly smaller values of k.

To be more precise, consider the case when ρ is a constant, independent of n. The lower bound of Theorem 1 implies that it is impossible to have $R_*^{\max}\to 0$ unless $k^2=o(e^{\rho m}/n)$. Thus, when the sample size m is so small that $m=o(\log n)$, then a vanishing risk is impossible to achieve unless $k\geq n^{1/2-o(1)}$. On the other hand, if $k\gg n^{1/2}$, then already for $m=1$, one has $R_*^{\max}\to 0$, see Arias-Castro et al. (2012). This shows the surprising fact that the difference between a sample of size $m=1$ and $m=o(\log n)$ is negligible and repeated observations do not help much.

However, when the sample size becomes logarithmic in n, the situation changes dramatically. In fact, when $m=\Omega((1/\rho)\log n)$, one has $R_*^{\max}\to 0$ *for all values of* $k\geq 2$, and such a vanishing risk is achieved by the localized squared-sum test described in Section 3.2.

This example reveals an interesting "phase transition" that occurs when the sample size becomes logarithmic in n.

1.3.1 Computational considerations

The "local" test that achieves near-optimal behavior in a large range of the parameters is a scan statistic that requires the computation of a maximum over all $\binom{n}{k}$ subsets of components of size k. In its naive implementation, this test is computationally intractable, unless k is very small. We also believe that computing this test is a fundamentally hard computational problem. We do not have a rigorous argument to prove such a hardness result but it is worth pointing out that the problem is quite similar, in spirit, to the notoriously difficult *hidden clique problem*, see Alon et al. (1999).

What performance can we achieve with limited computational power? Such questions of trade-off between statistical performance and computational complexity are at the heart of high-dimensional statistics and machine learning. We probe this question and describe a family of tests that balances detection performance and computational complexity.

In particular, in Section 4.4 we design a test that achieves near-optimal performance (similar to that of the scan statistic) and may be computed in polynomial-time in n when $m=O(\log n)$, ρ is a constant, and $k\sim n^a$ for some $a\in(0,1/2)$.

1.3.2 An application in the study of random geometric graphs

In Section 7 we apply the lower bound for the optimal risk in the clique model in a perhaps unexpected context and derive a new lower bound for the clique number of a high-dimensional random geometric graph. The setup is as follows.

Consider a random geometric graph on the unit sphere in dimension m. The graph has n vertices, each corresponding to a random point on the unit sphere. Two vertices are connected by an edge if the inner product of the corresponding points is positive. In a recent paper, Devroye et al. (2011) studied the clique number (i.e., the size of the largest clique in the graph) $\omega(n, m)$ of such a graph in various regimes. They showed that when $m \sim c \log n$ for a sufficiently small constant c, $\omega(n, m) = n^{1-o(1)}$ with high probability, while when $m \geq 9 \log^2 n$, $\omega(n, m) = O(\log^3 n)$. However, nothing was known about the behavior of the clique number in between. In particular, it was unclear where exactly the clique number becomes polylogarithmic. In Section 7 we show that the phase transition occurs at $m \asymp \log^2 n$. In particular, we prove that for all $c > 0$, when $m \sim c \log n$, then the median of $\omega(n, m)$ grows as a positive power of n and even for $m \asymp \log^{2-\epsilon} n$, the median of $\omega(n, m)$ grows faster than any power of $\log n$, for all $\epsilon > 0$.

1.4 Related work

As mentioned before, the literature on *sparse* covariance estimation and on graphical model estimation has become quite extensive. In spite of this surge of interest in sparse high-dimensional models, not much has been done in terms of detection of correlations. We note the work of Verzelen and Villers (2010), who consider the task of testing a given dependency structure. Our objective here is admittedly more modest and a more closely related is our own paper Arias-Castro et al. (2012), which focuses entirely on the case where the sample size is equal to one (i.e., $m = 1$). Our results here are seen to extend those in the one-sample case, with the regimes now partitioned according to the sample size. While the case where $\rho \to 1$ showed to be of interest in the case where $m = 1$, we are concerned here with the situation where $\rho \in (0, 1)$ is either fixed or tends to zero.

Note that our work is different from Butucea and Ingster (2011) where the task is the detection of a submatrix with higher per-coordinate mean in a large matrix with i.i.d. Gaussian entries, which is more closely related to the literature on the detection of sparse nonzero entries in the mean of a random vector. Our work has parallels with that literature which, for the clique model, focuses on the "detection-of-means" problem (see Addario-Berry et al. (2010); Arias-Castro et al. (2008); Baraud (2002); Donoho and Jin (2004); Hall and Jin (2010); Ingster (1999); Jin (2003)) defined as follows : Under the null, the vectors X_t are i.i.d. standard normal, while under the alternative, there is a subset $S \subset \{1, \dots, n\}$ in some class $\mathcal{C}$ of interest such that, X_t are i.i.d. normal with mean $(\mu_1, \dots, \mu_n)^T$ and identity covariance, where $\mu_i \geq \mu$ for $i \in S$ and $\mu_i = 0$ for $i \notin S$. Thus $\mu > 0$ is the minimum (per-coordinate) signal amplitude. Of course, one immediately reduces by sufficiency to the case $m = 1$ by averaging over the sample. This explains why the literature focuses on the case $m = 1$. The connection between the detection-of-means problem with the correlation detection problem studied here was detailed (for $m = 1$) in our previous paper (Arias-Castro et al., 2012), where ρ was found to correspond to μ^2. The connection is based on the following simple representation result for equi-correlated normal random variables.

Lemma 1 (**Berman (1962)**) *Let $X_1, \dots, X_k$ be standard normal random variables with $Cov(X_i, X_j) = \rho > 0$ for $i \neq j$. Then there are independent standard normal random variables $V, Y_1, \dots, Y_k$ such that $X_i = \sqrt{\rho}\, V + \sqrt{1-\rho}\, Y_i$ for all i.*

Thus, given V, the problem becomes that of detecting a subset of variables—here implicitly assumed to be indexed by $S = \{1, \dots, k\}$—with nonzero mean (equal to $\sqrt{\rho}\, V$) and with a variance equal to $1 - \rho$ (instead of 1). This representation was used in Arias-Castro et al. (2012) to obtain a general lower bound that seemed otherwise out of reach of more standard methods based on the second moment of the likelihood ratio.

This connection with the detection-of-means problem also applies in the case where $m > 1$, but with a twist. Indeed, when detecting correlations one does not average the vectors X_t but their covariances. So a simple reduction to the case $m = 1$ does not apply. However, one may still apply the representation result Lemma 1 to each observation vector X_t, yielding V_t's and $Y_{t,i}$'s that are independent standard normal random variables. By conditioning on $V_1, \dots, V_m$, the problem becomes equivalent to detecting a subset of variables with means $\sqrt{\rho}\, V_t$, $t = 1, \dots, m$. What makes the situation more complex is that the signs of the V_t's are random. Our approach to finding a general lower bound is based on this representation without which more standard methods seem to fail. The general lower bound, which is the key technical result of this paper, is given in Theorem 1 below.

1.5 Contribution and content of the paper

We obtain a general lower bound in Section 2 akin to, but not a straightforward extension of, the lower bound we obtained in Arias-Castro et al. (2012). We then study a number of tests that are near optimal in the sense that they come close to achieving the detection lower bound for various models. This is done in Section 3, where we also discuss computational issues, particularly in the clique model. We then specialize these general results in Sections 4, 5 and 6, to the three models described in Section 1.1. In Section 7, we apply our general lower bound to the problem of studying the size of the clique number of a random geometric graph on a high-dimensional sphere. We close the paper with a discussion in Section 8 of possible extensions and challenges.

2 Lower bounds

In this section we derive a general lower bound for the minimax risk $R_*^{\max}$. As mentioned in Section 1.2, the first step is to restrict the supremum in the definition of $R^{\max}(f)$ to covariance matrices in which all the nonzero entries are equal to $\rho > 0$ and then lower bound the maximum by an average. In particular, we have $R_*^{\max} \geq R^*$ where $R^* = \inf_f R(f)$ and

$$R(f) \stackrel{\text{def}}{=} \mathbb{P}_0\{f(X_1, \dots, X_m) = 1\} + \frac{1}{|\mathcal{C}|} \sum_{S \in \mathcal{C}} \mathbb{P}_S\{f(X_1, \dots, X_m) = 0\} \; .$$

Note that R^* is just the Bayes risk for the uniform prior on the models $S \in \mathcal{C}$. It is well known that the test f^* that achieves the infimum (i.e., $R(f^*) = R^*$) is the *likelihood ratio* *test* that accepts the null if and only if the likelihood ratio $(1/|\mathcal{C}|)\sum_{S\in\mathcal{C}}\phi_S(X)/\phi_0(X)<1$. Here ϕ_0 is the standard normal density in $\mathbb{R}^{nm}$ and ϕ_S is the normal density with covariance matrix defined by S. We focus on the case where ρ is bounded away from 1.

Theorem 1 *For any class $\mathcal{C}$, any $\rho\in[0,0.9)$, and any $a\geq\sqrt{8}$,*

$$R^*\geq \mathbf{P}\left\{\chi_m^2\leq ma^2\right\}\left(1-\frac{1}{2}\sqrt{\mathbb{E}\cosh^m(\nu_a Z)-1}\right)\ ,$$

where χ_m has chi-squared distribution with m degrees of freedom, $\nu_a:=\rho a^2/(1-\rho^2)$, and Z is the size of the intersection of two elements of $\mathcal{C}$ drawn independently and uniformly at random. In particular, taking $a=\sqrt{8}$,

$$R^*\geq\frac{1}{2}-\frac{1}{4}\sqrt{\mathbb{E}\cosh^m\left(\frac{8\rho}{1-\rho^2}Z\right)-1}\ ,$$

while $R^*\to 1$ if there is $a\to\infty$ such that $\mathbb{E}\cosh^m(\nu_a Z)\to 1$.

To appreciate the relative simplicity of the arguments that follow, we encourage the reader to try the bread-and-butter second moment method, which amounts to bounding the variance of the likelihood ratio directly. We avoid this by representing the data in a different way.

Proof. Under the alternative H_1, $X\in\mathbb{R}^{m\times n}$ can be written as

$$X_{t,i}=\begin{cases} Y_{t,i} & \text{if } i\notin S, t\in[m]\\ \sqrt{\rho}\,V_t+\sqrt{1-\rho}\,Y_{t,i} & \text{if } i\in S, t\in[m]\end{cases} \tag{2.1}$$

where $(Y_{t,i})_{i\in[n],t\in[m]}$, $(V_t)_{t\in[m]}$ are i.i.d. standard normal random variables.

Let $\varepsilon_t=\text{sign}(V_t)$, which are i.i.d. Rademacher, and $U_t=|V_t|$. Denote by U the m-vector with components $(U_1,\ldots,U_m)$. We consider now the alternative $H_1(u)$, defined as the alternative H_1 given $U=u\in\mathbb{R}^m$. Let $R(f)$, L, f^* (resp. $R_u(f)$, L_u, f_u^*) be the risk of a test f, the likelihood ratio, and the optimal (likelihood ratio) test, for H_0 versus H_1 (resp. H_0 versus $H_1(u)$). For any $u\in\mathbb{R}^m$, $R_u(f_u^*)\leq R_u(f^*)$, by the optimality of f_u^* for H_0 vs. $H_1(u)$. Therefore, conditioning on U,

$$R^*=R(f^*)=\mathbb{E}_U R_U(f^*)\geq\mathbb{E}_U R_U(f_U^*)=1-\frac{1}{2}\mathbb{E}_U\mathbb{E}_0|L_U(X)-1|\ .$$

($\mathbb{E}_U$ is the expectation with respect to U.) Using the fact that $\mathbb{E}_0|L_u(X)-1|\leq 2$ for all u, we have (with $B(0,a)$ being the euclidean ball of radius a in $\mathbb{R}^m$)

$$\mathbb{E}_U\mathbb{E}_0|L_U(X)-1|\leq 2\mathbb{P}\{||U||>a\sqrt{m}\}+\mathbb{P}\{||U||\leq a\sqrt{m}\}\max_{u\in B(0,a\sqrt{m})}\mathbb{E}_0|L_u(X)-1|\ .$$

Therefore, using the Cauchy-Schwarz inequality,

$$\begin{aligned}1-\frac{1}{2}\mathbb{E}_U\mathbb{E}_0|L_U(X)-1| &\geq \mathbb{P}\{||U||\leq a\sqrt{m}\}\left(1-\frac{1}{2}\max_{u\in B(0,a\sqrt{m})}\mathbb{E}_0|L_u(X)-1|\right)\\ &\geq \mathbb{P}\{||U||\leq a\sqrt{m}\}\left(1-\frac{1}{2}\max_{u\in B(0,a\sqrt{m})}\sqrt{\mathbb{E}_0 L_u^2(X)-1}\right)\ .\end{aligned}$$

We turn our attention to bounding $\mathbb{E}_0 L_u^2(X)$ from above. Let $L_{u,\varepsilon,S}(x)$ denote the likelihood ratio when S is anomalous, given u and ε, which is equal to

$$L_{u,\varepsilon,S}(x) \quad = \quad \frac{1}{(1-\rho)^{mk/2}} \exp\left(\sum_{t=1}^m \sum_{i\in S} \frac{x_{t,i}^2}{2} - \frac{(x_{t,i}-\sqrt{\rho}\varepsilon_t u_t)^2}{2(1-\rho)}\right).$$

Since $L_u(x) = \mathbb{E}_\varepsilon \mathbb{E}_S L_{u,\varepsilon,S}(x)$, by Fubini's theorem, we have

$$\mathbb{E}_0 L_u(X)^2 = \mathbb{E}_{S,S'} \mathbb{E}_{\varepsilon,\varepsilon'} \mathbb{E}_0 L_{u,\varepsilon,S}(X) L_{u,\varepsilon',S'}(X)$$

where $\varepsilon, \varepsilon'$ are i.i.d. Rademacher vectors and S, S' are i.i.d. uniform in the class $\mathcal{C}$. We have

$$L_{u,\varepsilon,S}(x) L_{u,\varepsilon',S'}(x) = (1-\rho)^{-mk} \exp(H_1(x) + H_2(x) + H_3(x)),$$

where

$$H_1(x) := \sum_{t=1}^m \sum_{i\in S\cap S'} x_{t,i}^2 - \frac{(x_{t,i}-\sqrt{\rho}\varepsilon_t u_t)^2}{2(1-\rho)} - \frac{(x_{t,i}-\sqrt{\rho}\varepsilon_t' u_t)^2}{2(1-\rho)},$$

$$H_2(x) := \sum_{t=1}^m \sum_{i\in S\setminus S'} \frac{x_{t,i}^2}{2} - \frac{(x_{t,i}-\sqrt{\rho}\varepsilon_t u_t)^2}{2(1-\rho)},$$

$$H_3(x) := \sum_{t=1}^m \sum_{i\in S'\setminus S} \frac{x_{t,i}^2}{2} - \frac{(x_{t,i}-\sqrt{\rho}\varepsilon_t' u_t)^2}{2(1-\rho)} \; .$$

Let $Z = |S \cap S'|$. We see that $H_1(X), H_2(X), H_3(X)$ are independent of each other under the null with

$$\mathbb{E}_0 \exp(H_2(X)) = (1-\rho)^{m|S\setminus S'|/2} = (1-\rho)^{m(k-Z)/2}$$

and similarly for $\mathbb{E}_0 \exp(H_3(X))$, while

$$\mathbb{E}_0 \exp(H_1(X)) = \left(\frac{1-\rho}{1+\rho}\right)^{mZ/2} \exp\left(\frac{\rho Z}{1-\rho^2} \sum_{t=1}^m \varepsilon_t \varepsilon_t' u_t^2 - \frac{\rho^2 Z}{1-\rho^2} \|u\|^2\right).$$

For the latter, we used the fact that $\varepsilon_t^2 = {\varepsilon_t'}^2 = 1$, to get

$$\begin{aligned}
&\int_{\mathbb{R}} \exp\left(x_{t,i}^2 - \frac{(x_{t,i}-\sqrt{\rho}\varepsilon_t u_t)^2}{2(1-\rho)} - \frac{(x_{t,i}-\sqrt{\rho}\varepsilon_t' u_t)^2}{2(1-\rho)}\right) \exp(-x_{t,i}^2/2) \frac{dx_{t,i}}{\sqrt{2\pi}} \\
= \; &\int_{\mathbb{R}} \exp\left(\frac{\rho(\varepsilon_t \varepsilon_t' - \rho) u_t^2}{1-\rho^2} - \frac{1+\rho}{2(1-\rho)} \left(x_{t,i} - \frac{\sqrt{\rho}}{1+\rho}(\varepsilon_t + \varepsilon_t') u_t\right)^2\right) \frac{dx_{t,i}}{\sqrt{2\pi}} \\
= \; &\sqrt{\frac{1-\rho}{1+\rho}} \exp\left(\frac{\rho(\varepsilon_t \varepsilon_t' - \rho) u_t^2}{1-\rho^2}\right) \; ,
\end{aligned}$$

where the last line comes from a simple change of variables. Hence,

$$\mathbb{E}_0 L_{u,\varepsilon,S}(x) L_{u,\varepsilon',S'}(X) = (1-\rho^2)^{-mZ/2} \exp\left(\frac{\rho Z}{1-\rho^2} \sum_{t=1}^m \varepsilon_t \varepsilon_t' u_t^2 - \frac{\rho^2 Z}{1-\rho^2} \|u\|^2\right).$$

Since $(\varepsilon_t \varepsilon'_t : t = 1, \dots m)$ are i.i.d. Rademacher, we have

$$\mathbb{E}_{\varepsilon,\varepsilon'} \exp\left(\frac{\rho Z}{1-\rho^2} \sum_{t=1}^m \varepsilon_t \varepsilon'_t u_t^2\right) = \prod_{t=1}^m \cosh\left(\frac{\rho Z u_t^2}{1-\rho^2}\right).$$

Define

$$\gamma = \frac{\rho Z}{1-\rho^2}, \quad b_t = \gamma^{1/2} u_t.$$

Maximizing

$$\prod_{t=1}^m \cosh(b_t^2) \exp(-\rho b_t^2) \quad \text{subject to} \quad \sum_{t=1}^m b_t^2 \le m\gamma a^2 \ , \tag{2.2}$$

using Lagrangian multipliers and checking the Karush-Kuhn-Tucker conditions, we find that at a local maximum all the b_t must be equal, so that the value of the optimization problem (2.2) is equal to

$$\left(\max_{0 \le c \le \gamma a^2} \cosh(c) \exp(-\rho c)\right)^m . \tag{2.3}$$

The function $c \mapsto \cosh(c)\exp(-\rho c)$ is decreasing on $(0, \rho)$ and increasing on (ρ, ∞). Therefore the maximum is either at $c = 0$ or $c = \gamma a^2$, and comparing the two, we have

$$\cosh(c)\exp(-\rho c) = 1 \Leftrightarrow g(c) := \frac{1}{c} \log\cosh(c) = \rho \ .$$

Straightforward calculations show that g is strictly increasing on $(0, \infty)$ with range $(0, 1)$. Therefore $w(\rho) = g^{-1}(\rho)$ is well-defined and, as a function of ρ, is infinitely differentiable and increasing. Hence, the maximum of (2.3) is at $c = \gamma a^2$ if $w(\rho) < \gamma a^2$. Again, elementary calculations show that

$$w' = \frac{w}{\tanh(w) - \rho}, \ w'' = \frac{w \tanh^2(w)}{(\tanh(w) - \rho)^2},$$

implying in particular that w is convex. Numerically we find that $w(0.9) < 7$, so that (by convexity) $w(\rho) \le 8\rho$ for all $\rho \le 0.9$, and $w(\rho) < \gamma a^2$ when $a \ge \sqrt{8}$ and $Z \ge 1$.

Since we assume that $a \ge \sqrt{8}$ and $\rho \le 0.9$, (2.3) is equal to

$$\cosh^m(\gamma a^2) \exp(-m\rho\gamma a^2).$$

(This is also true when $Z = 0$.) Hence,

$$\begin{aligned}
&\mathbb{E}_{\varepsilon,\varepsilon'} \mathbb{E}_0 L_{u,\varepsilon,S}(x) L_{u',\varepsilon',S'}(X) \\
&= (1-\rho^2)^{-mZ/2} \prod_{t=1}^m \cosh(b_t^2) \exp(-\rho b_t^2) \\
&\le \cosh^m\left(\frac{\rho a^2 Z}{1-\rho^2}\right) \exp\left(-\frac{m\rho^2 a^2 Z}{1-\rho^2} - \frac{mZ}{2} \log(1-\rho^2)\right) \\
&\le \cosh^m\left(\frac{\rho a^2 Z}{1-\rho^2}\right),
\end{aligned}$$

where in the last inequality we used the fact that $a \geq 1$ and $s + \frac{1-s}{2}\log(1-s) \geq 0$ for all $s \in (0,1)$. $\square$

In Sections 4, 5, and 6, we specialize Theorem 1 to the different models we described in Section 1.1. Throughout, we assume that ρ is bounded away from 1. Specifically, we fix $\rho_0 < 1$ and consider $\rho \leq \rho_0$ (in the proof above we chose $\rho_0 = 0.9$, but the proof technique works for any $\rho_0 < 1$). We also assume that $k/n \to 0$. When $k \asymp n$, then $R^* \to 1$ if $\rho n\sqrt{m} \to 0$. This is a straightforward application of our result using the bound $Z \leq n$. We leave the (easy) details to the reader.

3 Tests

In this section we introduce and briefly discuss two natural tests that will be seen to perform near optimally in various regimes of the parameters. This optimality property will be established in Sections 4, 5, and 6, by comparing simple performance bounds with the implications of Theorem 1.

The first test, that we call "squared-sum test", is based on a global test statistic that does not take the class $\mathcal{C}$ into account at all.

The second test, a "localized" squared-sum test, is based on a simple scan statistic. It may also be interpreted as a simplified version of the generalized likelihood ratio test.

As we will see, one of the two tests above always has a near-optimal performance in all three specific classes we discuss. Thus, the story is essentially complete for the point of view of detection performance. Unfortunately, the localized squared-sum test is hard to compute. We discuss two possible substitutes. The first one is a simple "maximum correlation test" that turns out to be nearly optimal for very small values of k. In Section 4.4 we discuss another test in the context of the clique model that is both near-optimal and computationally feasible when the sample size is at most logarithmic in the dimension n.

All performance bounds derived below are in terms of the average correlation

$$\rho_{\text{ave}} = \frac{1}{k(k-1)} \sum_{i,j \in S : i \neq j} \sigma_{i,j} \geq \rho \ , \tag{3.1}$$

where S is the anomalous set.

3.1 The squared-sum test

Consider the *squared-sum test* that rejects for large values of the test statistic

$$Y = \sum_{t=1}^{m} \left(\sum_{i=1}^{n} X_{t,i} \right)^2 \ . \tag{3.2}$$

Such a test relies on the fact that all correlations are non-negative under the alternative hypothesis. The following result gives a simple characterization of the performance of the squared-sum test. Since the test does not use information about the class $\mathcal{C}$, its minimax risk does not depend on the model either.

Proposition 1 *The squared-sum test that rejects H_0 when $Y \geq nm + m^{3/4} k \sqrt{n\rho_{\text{ave}}}$ is asymptotically powerful when $\rho_{\text{ave}} \sqrt{m}\, k^2/n \to \infty$. The squared-sum test with any threshold value for Y is asymptotically powerless when $\rho_{\text{ave}} \sqrt{m}\, k^2/n \to 0$.*

Proof. Suppose that $a := \rho_{\text{ave}} \sqrt{m}\, k^2/n \to \infty$. Under the null $Y \sim n\chi^2_m$, so that

$$\mathbb{P}_0(Y > n(m + \sqrt{am})) \to 0, \quad n \to \infty.$$

Under the alternative $Y \sim (n + \rho_{\text{ave}}\, k(k-1))\chi^2_m$, so that

$$\mathbb{P}_1(Y \leq (n + \rho_{\text{ave}}\, k(k-1))(m - \sqrt{am})) \to 0, \quad n \to \infty.$$

Since

$$(n + \rho_{\text{ave}}\, k(k-1))(m - \sqrt{am}) - n(m + \sqrt{am}) = n\sqrt{m}(a + o(a)) \to \infty,$$

the test with critical value $n(m + \sqrt{am})$ is asymptotically powerful.

Suppose that $a \to 0$. We still have that $Y/n \sim \chi^2_m$ under H_0 while $Y/n \sim (1 + \rho_{\text{ave}}\, k(k-1)/n)\chi^2_m$ under H_1. If m is fixed, Y/n is asymptotically χ^2_m under the alternative since $\rho_{\text{ave}}\, k(k-1)/n \to 0$ in that case. If $m \to \infty$, $(Y/n - m)/\sqrt{2m}$ is asymptotically standard normal both under the null and the alternative. The latter is due to Slutsky's Theorem, since

$$\frac{Y/n - m}{\sqrt{2m}} = \frac{\chi^2_m - m}{\sqrt{2m}} + \frac{\rho_{\text{ave}}\, k(k-1)\chi^2_m}{\sqrt{2m}} = \frac{\chi^2_m - m}{\sqrt{2m}} + O_P(a),$$

using the fact that $\chi^2_m = O_P(m)$. □

3.2 A localized squared-sum test

When k is smaller, global tests such as the squared-sum test are not very powerful. The generalized likelihood ratio test "scans" over all subsets S in the class $\mathcal{C}$. Instead of studying the generalized likelihood ratio test, we consider a localized version of the squared-sum test that has similar power and is a little easier to analyze. The *localized squared-sum test* rejects for large values of the test statistic

$$Y_{\text{scan}} = \max_{S \in \mathcal{C}} Y_S, \quad \text{where} \quad Y_S = \sum_{t=1}^{m} \left(\sum_{i \in S} X_{t,i} \right)^2 .$$

The following result gives sufficient conditions for the test to be asymptotically powerful. The conditions are in terms of the cardinality of the class $\mathcal{C}$. Sharper bounds that take into account the fine metric structure of $\mathcal{C}$ are also possible by more careful bounding of the distribution of Y_{scan} under the null. However, as we will see below, this bound is already quite sharp for the specific classes considered in this paper, and to preserve relative simplicity of the arguments we do not pursue sharper bounds here.

Proposition 2 *The localized squared-sum test that rejects the null hypothesis if*

$$Y_{\text{scan}} > \frac{1}{2} \left(\rho_{\text{ave}} k^2 m + \left(H^{-1}(3 \log |\mathcal{C}|/m) - 1 \right) km \right),$$

is asymptotically powerful when

$$\rho_{\text{ave}} k \geq 2H^{-1}(3\log(|\mathcal{C}|)/m),$$

where $H(b) := b - 1 - \log b$ *for* $b > 1$.

Proof. Under the null $Y_S \sim k\chi^2_m$ for all $S \in \mathcal{C}$, with

$$\mathbf{P}\left\{\chi^2_m > bm\right\} \leq \exp[-(m/2)(b - 1 - \log b)], \quad \forall b \geq 1, \tag{3.3}$$

by Chernoff's bound. Hence, with the union bound,

$$\mathbb{P}_0(Y_{\text{scan}} > bkm) \leq |\mathcal{C}| \exp[-(m/2)(b - 1 - \log b)].$$

By letting $b = H^{-1}(3\log|\mathcal{C}|/m)$, the probability above tends to zero.

Under the alternative where S is anomalous, $Y_S \sim (k + \rho_{\text{ave}}\, k(k-1))\chi^2_m$, so that

$$Y_{\text{scan}} \geq Y_S > km + \rho_{\text{ave}}\, k^2 m - O_P((k + \rho_{\text{ave}}\, k^2)\sqrt{m}).$$

Hence, the test is asymptotically powerful when $\rho_{\text{ave}}\, k^2 m \geq 2km\,(H^{-1}(3\log|\mathcal{C}|/m) - 1)$. □

Note that when $b \to 1$, we have $H(b) \sim (b-1)^2/2$ and therefore in the case when $(\log|\mathcal{C}|)/m \to 0$, the test is asymptotically powerful for $\rho_{\text{ave}} k \geq A\sqrt{(\log|\mathcal{C}|)/m}$ for a constant A sufficiently large. On the other hand, $H(b) \sim b$ when $b \to \infty$, so in the case when $(\log|\mathcal{C}|)/m \to \infty$, the sufficient condition for ρ_{ave} is that $\rho_{\text{ave}} k \geq A(\log|\mathcal{C}|)/m$. Put it another way, a sufficient condition for the test to be asymptotically powerful is that

$$\rho_{\text{ave}} k \geq A \max\left(\sqrt{(\log|\mathcal{C}|)/m},\ (\log|\mathcal{C}|)/m\right) .$$

When the class $\mathcal{C}$ is large (i.e., has size exponential in k), the test statistic may be difficult to compute as it involves solving a nontrivial combinatorial optimization problem. This is the case for the clique model (unless k is very small) and the matching model. In fact, we believe that the problem of computing, or even approximating, Y_{scan} is fundamentally hard, though we do not have a formal argument to prove it. An even stronger conjecture is that, for the clique model, computing *any* test with a near-optimal performance is fundamentally hard in some range of the parameters. We believe this is a challenging and important research problem. In Section 3.3 we suggest a test that has good performance and that is efficiently computable if m is only logarithmic in n.

3.3 Maximum correlation test

Finally, we mention the possibly simplest test that one would think of when confronted with testing H_0 in the sparse regime. This is the test that rejects for large values of the maximum pairwise empirical correlation

$$Y_{\max} = \max_{i \neq j} \sum_{t=1}^{m} X_{t,i} X_{t,j} \ .$$

In fact, this test does have some power in the sparse regime, and is actually near-optimal when k is fixed as the following result shows. However, one cannot expect a good performance of this test for large values of k. An advantage of this test is that it may be computed efficiently in a straighforward manner.

Corollary 1 *The maximum correlation test tat rejects* H_0 *when* $Y_{\max} > \sqrt{5m \log n}$ *is asymptotically powerful when*

$$\rho_{\text{ave}} \geq \sqrt{5(\log n)/m} \ .$$

Proof. Assume that $m \geq 5 \log n$ for otherwise the statement is void. For $i \neq j$ fixed, under the null, $X_{t,i}X_{t,j}, t = 1, \ldots, m$ are i.i.d. with zero mean, unit variance, and finite moment generating function in the a neighborhood of the origin. In fact, it is equal to $(1 - \lambda^2)^{-1/2}$ for $\lambda \in (-1, 1)$. Hence, by a standard result on moderate deviations (Dembo and Zeitouni, 2010, Th 3.7.1),

$$\limsup_{m \to \infty} \frac{m}{b_m^2} \log \mathbb{P}_0 \left\{ \sum_{t=1}^{m} X_{t,i}X_{t,j} > b_m \right\} \leq -\frac{1}{2}$$

for any sequence (b_m) such that $\sqrt{m} \ll b_m \ll m$. We choose $b_m = \sqrt{5m \log n}$ and use the union bound, to get

$$\mathbb{P}_0\{Y_{\max} > \sqrt{5m \log n}\} \leq \binom{n}{2} \mathbb{P}_0 \left\{ \sum_{t=1}^{m} X_{t,i}X_{t,j} > b_m \right\} \to 0 \ .$$

Under the alternative when $S \subset [n]$ is anomalous, pick $i \neq j$ in S such that $X_{t,i}X_{t,j}, t = 1, \ldots, m$ are i.i.d. with mean larger than ρ_{ave} and variance smaller than 2, so that by Chebyshev's inequality,

$$\sum_{t=1}^{m} X_{t,i}X_{t,j} = m\rho_{\text{ave}} + O_P(\sqrt{m}) \ .$$

From this, the result follows immediately. □

Note that the maximum correlation test is more powerful than the squared-sum test when

$$\frac{k^2}{n}\sqrt{\log n} \to 0.$$

4 Clique model

In this section we focus on the clique model. We derive corollaries of the Theorem 1 in various ranges of the parameters and compare them with the performance bounds for the squared-sum test and the scan statistics-based test considered in Section 3.

4.1 Lower bounds for the clique model

In order to apply Theorem 1, note that in the clique model, Z is hypergeometric with parameters (n, k, k), which is stochastically bounded by the binomial distribution with parameters $(k, k/(n - k))$.

Recall that by assumption, there exists a $\rho_0 < 1$ such that for all n, $\rho \leq \rho_0$. We distinguish various regimes of the parameters in which the minimax risk and the optimal test behave differently.

- **Case 1.** Suppose

$$\frac{k}{n} \to 0 \quad \text{and} \quad \frac{k^2}{n} \to \infty \quad \text{and} \quad \rho\sqrt{m}\,\frac{k^2}{n} \to 0 \quad \text{and} \quad \begin{cases} \text{either } \rho m \to 0\ , \\ \text{or } \rho\sqrt{mk} \to 0\ . \end{cases} \tag{4.1}$$

Let $\zeta = \rho\sqrt{m}k^2/n$ and choose $a \to \infty$ such that $a^2\zeta \to 0$. When $Z \le 8k^2/n$, we use the fact that $\rho a^2 Z \to 0$ and $\cosh(x) = 1 + x^2/2 + o(x^2)$ when $x \to 0$ to get that for all sufficiently large n,

$$\cosh^m(\nu_a Z) = \big(1 + (\nu_a Z)^2/2 + o(\nu_a Z)^2\big)^m \le \exp(64 C_0^2 a^4 \zeta^2) = 1 + o(1)$$

where $C_0 = (1-\rho_0^2)^{-1}$. So it suffices to show that

$$\mathbb{E}\left[\cosh^m(\nu_a Z)\mathbb{1}_{\{Z > 8k^2/n\}}\right] = o(1).$$

If $\rho m \to 0$, we choose a such that $a^2\rho m \to 0$. We use the bound $\cosh(x) \le \exp(x)$ and Bennett's inequality, to get

$$\begin{aligned} \mathbb{E}\left[\cosh^m(\nu_a Z)\mathbb{1}_{\{Z > 8k^2/n\}}\right] &\le \sum_{z > 8k^2/n} \exp\big(-z(1 - C_0 a^2 \rho m)\big) \\ &\le \exp\big(-7k^2/n\big) \to 0, \end{aligned}$$

eventually.

If $\rho\sqrt{mk} \to 0$, we may assume that $k \le m$ for otherwise $\rho m \to 0$, which we already covered. We choose a such that $a^2\rho\sqrt{mk} \to 0$ (which implies in particular $a^2\rho k \to 0$). We use the bounds $\cosh(x) \le 1 + x^2/2 + o(x^2)$ and $Z \le k$, and the fact that $a^2\rho k \to 0$, to get

$$\cosh^m(\nu_a Z) \le \exp(C_0^2 a^4 \rho^2 m Z^2) \le \exp(C_0^2 a^4 \rho^2 mkZ),$$

eventually. Combined with Bennett's inequality, we get

$$\begin{aligned} \mathbb{E}\left[\cosh^m(\nu_a Z)\{Z > 8k^2/n\}\right] &\le \sum_{z > 8k^2/n} \exp\big(-z(1 - C_0^2 a^4 \rho^2 mk)\big) \\ &\le \exp\big(-7k^2/n\big) \to 0, \end{aligned}$$

eventually.

- **Case 2.** Suppose

$$\frac{k^2}{n} \to 0 \quad \text{and} \quad \rho\sqrt{\frac{mk}{\log(n/k^2)}} \to 0 \quad \text{and} \quad \frac{k\log(n/k^2)}{m} \to 0. \tag{4.2}$$

Let $\zeta = \rho\sqrt{mk/\log(n/k^2)}$ and choose $a \to \infty$ such that $a\zeta \to 0$ and $a^2\rho k \to 0$. The latter is possible because (4.2) implies that $\rho k \to 0$. Then

$$\cosh^m(\nu_a Z) = (1 + (\nu_a Z)^2/2 + o(\nu_a Z)^2)^m \le \exp\big(C_0^2 a^4 \rho m Z^2\big),$$

eventually, where we used the bound $Z \leq k$ and $\nu_a Z = O(a^2 \rho k) = o(1)$. We then bound $Z^2 \leq kZ$ and use the fact that Z is stochastically bounded by $\text{Bin}(k, k/(n-k))$, and knowing the moment generating function of the latter, we have

$$\begin{aligned}
\mathbb{E}\cosh^m(\nu_a Z) &\leq \left(1 + \frac{k}{n-k} e^{C_0^2 a^4 m \rho^2 k}\right)^k \\
&\leq \exp\big(\exp\big(-\log(n/k^2)(1 - C_0^2 a^4 \zeta^2)\big)\big) = 1 + o(1)\ .
\end{aligned}$$

- **Case 3.** Suppose
$$\frac{k^2}{n} \to 0 \quad \text{and} \quad \rho \frac{m}{\log(n/k^2)} \to 0. \tag{4.3}$$
Let $\zeta = \rho m / \log(n/k^2)$ and choose $a \to \infty$ such that $a^2 \zeta \to 0$. We use the fact that $\cosh(x) \leq \exp(x)$, and use the same bound on the moment generating function of Z, to get (eventually)
$$\begin{aligned}
\mathbb{E}\cosh^m(\nu_a Z) &\leq \mathbb{E}\exp(m \nu_a Z) \\
&\leq \left(1 + \frac{k}{n-k} e^{C_0 a^2 \rho m}\right)^k \\
&\leq \exp\big(\exp\big(-\log(n/k^2)(1 - C_0^2 a^4 \zeta^2)\big)\big) = 1 + o(1).
\end{aligned}$$

This leads to the following.

Corollary 2 *In the clique model, under either* (4.1), (4.2), *or* (4.3), $R^* \to 1$.

We will see below that (4.1) is tight up to logarithmic factors. This is also the case of (4.2) and (4.3), unless $k^2/n \to 0$ as a negative power of n. Note also that the result is silent in the regime when $k^2/n \to 0$ and $\rho\sqrt{m}\, k^2/n \to 0$. However, it is covered by (4.2) when $\log(n/k^2)/k \to 0$, and by (4.3) when $\log(n/k^2)/\sqrt{m} \to 0$, so again it is a matter of logarithmic factors. We mention that the typical exposition in the detection-of-means literature, for example in Donoho and Jin (2004), avoids the discussion of such fine details by assuming that $k = n^\alpha$ for some $\alpha \in (0, 1)$.

4.2 Localized squared-sum test

Next we take a closer look at the performance of the localized squared-sum test for the clique model. In this case we have $|\mathcal{C}| = \binom{n}{k}$ so $\log|\mathcal{C}| \sim k \log(n/k)$. Plugging this into Proposition 2, we see that the local squared-sum test is reliable when

$$\rho_{\text{ave}} \geq (2/k) H^{-1}(3k \log(n/k)/m)\ .$$

Based on this and Corollary 2, we conclude that the test is near-optimal in regimes (4.2) and (4.3), though only up to a logarithmic factor if $k^2/n \to 0$ slower than any power of n.

However, we do not have such a guarantee in the regime (4.1). In this range of parameters, it is the squared-sum test that yields an optimal performance (up to a logarithmic factor) Also, comparing Proposition 1 and Proposition 2, we see that the local test dominates when $\max\big(1, (k/m)^{1/2}\big)\, k^{3/2}/n$ tends to zero faster than $1/\log(n/k)$.

4.3 The case of constant ρ

Now we discuss the simple but interesting case when ρ is bounded away from zero. For simplicity, we may assume that ρ is a constant, independent of N.

From our previous work Arias-Castro et al. (2012) (and also from Theorem 1) in the case of $m = 1$, we know that when $\rho < 1$ is constant, $R^* \to 1$ unless $k^2/n \to \infty$. Now we learn from Corollary 2 that $R^* \to 1$ also when $k^2/n \to 0$ and $m = o(\log(n/k^2))$. Hence, in the case of ρ constant and n/k^2 a positive power of n, a sample size m sub-logarithmic in the dimension n is not enough for reliable detection, and is qualitatively on par with the case of $m = 1$.

However, the situation dramatically changes when the sample size becomes at least logarithmic in the dimension n. Indeed, even for $k = 2$, both the localized squared-sum test and the maximum correlation test have a vanishing risk for any constant value of ρ when $(\log n)/m \to \infty$. This reveals a very interesting "phase transition" occuring when the sample size is about logarithmic in the dimension.

4.4 Balancing detection and running time

Given the often enormous size of data sets that statisticians need to handle as an every-day practice, it is of great interest to design computationally efficient, yet near-optimal tests. In the case of the clique model, this is a highly non-trivial task, because the class $\mathcal{C}$ has size exponential in k and therefore computing the localized squared-sum test (or other versions of the generalized likelihood ratio test and scan statistics) involves a non-trivial optimization problem over all $\binom{n}{k}$ elements of $\mathcal{C}$. In fact, often it seems that small testing risk and computational efficiency are contradicting terms. In this section we show that in at least one non-trivial instance, it is possible to design a computationally efficient (i.e., computable in time quadratic in n) test that has near optimal risk.

This is the case when the sample size m is (at most) logarithmic in n and $k \sim n^a$ for some $a \in (0, 1)$. (Recall from Section 4.3 that this is a quite interesting range of parameters.)

To introduce a family of tests that balance detection performance and computational complexity, let $\ell \in \{1, \ldots, m\}$ and define

$$Y(\ell) = \max_{S:|S|=k} \max_{T:|T|=\ell} \sum_{t \in T} \sum_{i \in S} X_{t,i} \ .$$

Since

$$Y(\ell) = \max_{T:|T|=\ell} \sum_{t \in T} \sum_{i=n-k+1}^{n} X_{t,(i)} \ ,$$

where $X_{t,(1)} \leq \cdots \leq X_{t,(n)}$ are the ordered $X_{t,i}$'s, the statistic $Y(\ell)$ can be computed in $O(\binom{m}{\ell}(n \log(n) k + \ell \log(m)))$ time by first sorting $(X_{t,i} : i = 1, \ldots, n)$ and summing the largest k, for all subsets T of size ℓ, and then maximizing over these.

For example, when $m \leq \log_2 n$, then $\binom{m}{\ell} \leq 2^m \leq n$ and the test may be computed in time $O(n^2 \log n)$. Even when $m \sim C \log n$ for some constant $C > 0$, we may choose $\ell \sim \gamma m$ such that $C\gamma \log(1/\gamma) \leq 1$. In that case $\binom{m}{\ell} \leq 2^{\ell \log_2(m/\ell)} \leq n$ and again the test may be computed in time $O(n^2 \log n)$. The next proposition bounds the risk of the test.

Proposition 3 *Then the test based on $Y(\ell)$ with $\ell \le m/7$ is asymptotically powerful in the clique model when*

$$\rho_{\text{ave}} \ge 3 \left(\frac{\log(n/k)}{\ell} + \frac{\log(m/\ell)}{k} \right) .$$

Proof. Since under the null $\sum_{t\in T}\sum_{i\in S} X_{t,i} \sim \mathcal{N}(0, \ell k)$, by a standard bound for the maximum of a finite set of Gaussian variables,

$$Y(\ell) \le \sqrt{2\ell k \log \binom{m}{\ell}\binom{n}{k}} \sim \sqrt{2\ell k(\ell \log(m/\ell) + k\log(n/k))}$$

with high probability. Under the alternative where S is anomalous, we have

$$Y(\ell) \ge \sqrt{\rho_{\text{ave}}}\, k \left(Z_{(m-\ell+1)} + \cdots + Z_{(m)} \right) ,$$

where $Z_{(1)} \le \cdots \le Z_{(m)}$ are the ordered values of

$$Z_t := (k + \rho_{\text{ave}}\, k(k-1))^{-1/2} \sum_{i\in S} X_{t,i} ,$$

which are i.i.d. standard normal. Since we assume that $m \to \infty$, we have that $\mathbb{P}(Z_{(m-\ell+1)} \ge 1) \to 1$ when $\ell/m \le 1/7 < \mathbf{P}\{\mathcal{N}(0,1) > 1\}$. Hence, $Y(\ell) \ge \sqrt{\rho_{\text{ave}}}\, k\ell$ with probability tending to one under the alternative. From this the result follows. $\square$

In the regime of (4.3) with $m \sim C \log n$, we see that the test is optimal up to a constant factor in ρ when $k \sim n^a$ for some $a < 1/2$. In this range of parameters, it seems hopeless to compute (or even approximate) the local squared-sum test.

However, when m is much larger than logarithmic in n, this test also requires super-polynomial computational time and therefore it is not useful in practice. In such cases one may have to resort to sub-optimal tests such as the maximum correlation test described in Section 3.3. It is an important and difficult challenge to find out the possibilities and limitations of powerful detection taking computational constraints into account.

5 Block model

Next we discuss the consequences of our main results for the block model which serves as an easy and prototypical example of "small" or "parametric" classes.

In this model, to apply Theorem 1, we may use the obvious bound $Z \le \tilde{Z} := k\mathbb{1}\{S \cap S' \ne \emptyset\}$. Noting that $\mathbb{P}(S \cap S' \ne \emptyset) \le 2k/n$, we have

$$\mathbb{E}\cosh^m(\nu_a Z) \le 1 + \frac{2k}{n}\cosh^m(\nu_a k) .$$

We distinguish between two main regimes.

- **Case 1.** Suppose

$$\rho k \sqrt{\frac{m}{\log(n/k)}} \to 0 \quad \text{and} \quad \frac{k}{n} \to 0 \quad \text{and} \quad \frac{\log(n/k)}{m} \to 0 \ . \tag{5.1}$$

Let $\zeta = \rho k \sqrt{m/\log(n/k)}$ and choose $a \to \infty$ such that $a^2\zeta \to 0$ and $a^2 \rho k \to 0$. The latter is possible because (5.1) implies that $\rho k \to 0$. We use the bound $\cosh(x) = 1 + x^2/2 + o(x^2)$ to get

$$\cosh^m(k\nu_a) = \big(1 + (k\nu_a)^2/2 + o(k\nu_a)^2\big)^m \le \exp(C_0^2 a^4 \rho^2 k^2 m) \ ,$$

for n sufficiently large. Then

$$\frac{2k}{n} \exp(C_0^2 a^4 \rho^2 k^2 m) = 2\exp(-\log(n/k)(1 - C_0^2 a^4 \zeta^2)) \to 0 \ ,$$

by our assumptions.

- **Case 2.** Suppose

$$\rho k \frac{m}{\log(n/k)} \to 0 \quad \text{and} \quad \frac{k}{n} \to 0 \ . \tag{5.2}$$

Let $\zeta = \rho k m/\log(n/k)$ and choose $a \to \infty$ such that $a^2\zeta \to 0$. We use the bound $\cosh(x) \le \exp(x)$ to get

$$\frac{2k}{n} \cosh^m(k\nu_a) \le 2\exp\big(-\log(n/k)(1 - C_0 a^2 \zeta)\big) \to 0 \ .$$

This leads to the following.

Corollary 3 *In the block model, under either* (5.1) *or* (5.2), $R^* \to 1$.

In view of Corollary 3, the squared-sum test is near-optimal for the block model only when $k \asymp n$. However, the localized squared-sum test has a much better performance. We have $|\mathcal{C}| = n$, and plugging this into Proposition 2, we see that the localized squared-sum test is reliable when

$$\rho_{\text{ave}} k \ge 2H^{-1}(3(\log n)/m) \ .$$

With Corollary 2, we conclude that the test is near-optimal except in the case where $k/n \to 0$ slower than any negative power of n, where the test is optimal up to a logarithmic factor.

6 Perfect matching model

Here we work out the corollaries of our main results for the perfect matching model. This model illustrates how one may proceed when the model in question has a non-trivial combinatorial structure. In order to use Theorem 1, one needs to use the specific properties of the class.

In the perfect matching model, Z is distributed as the number of fixed points in a random permutation over $\{1,\ldots,k\}$. It is well known that

$$\mathbb{P}\{Z=z\} = \frac{1}{z!}\sum_{s=0}^{k-z}\frac{(-1)^s}{s!} \le \frac{1}{z!}\left(\frac{1}{e}+\frac{1}{(k-z+1)!}\right) \ , \quad \forall z\in\{0,\ldots,k\}. \tag{6.1}$$

We distinguish between two main regimes. To simplify notation, we assume that k is even.

- **Case 1.** Suppose
$$\rho\sqrt{k\max(k,m)} \to 0 \ . \tag{6.2}$$
We choose $a\to\infty$ such that $a^2\rho\sqrt{k\max(k,m)}\to 0$. We use the bounds $\cosh(x) = 1+x^2/2+o(x^2)$ and $Z\le k$, and the fact that $a^2\rho k\to 0$, to get, for n sufficiently large,
$$\cosh^m(\nu_a Z) \le \exp(C_0^2 a^4\rho^2 m Z^2) \le \exp(C_0^2 a^4 \rho^2 mkZ).$$
Now let $c = C_0^2 a^4\rho^2 mk$. Using (6.1), one obtains
$$\begin{aligned}\mathbb{E}\cosh^m(\nu_a Z) &\le \mathbb{E}\exp(cZ)\\ &\le \sum_{z=0}^{k}\frac{1}{z!}\left(\frac{1}{e}+\frac{1}{(k-z+1)!}\right)\exp(cz)\\ &\le \exp(\exp(c)-1)+\frac{k+1}{(k/2+1)!}\exp(ck)\\ &\le 1+o(1),\end{aligned}$$
because $c\to 0$ and $\log[(k/2+1)!]\sim(k/2)\log k$ as $k\to\infty$.

- **Case 2.** Suppose
$$\frac{\rho m}{\log(\min(k,m))}\to 0. \tag{6.3}$$
We choose $a\to\infty$ such that $a^2\rho m/\log(\min(k,m))\to 0$. Using (6.1) one obtains
$$\begin{aligned}\mathbb{E}\cosh^m(\nu_a Z) &\le \mathbb{E}\cosh^m(\nu_a Z)\mathbb{1}_{\{Z<k/2\}} + \mathbb{P}\{Z\ge k/2\}\cosh^m(\nu_a k)\\ &\le \sum_{z=0}^{k/2-1}\frac{1}{z!}\left(\frac{1}{e}+\frac{1}{(k/2)!}\right)\cosh^m(\nu_a z)+\mathbb{P}\{Z\ge k/2\}\exp(\nu_a mk)\\ &\le \frac{1}{e}\sum_{z=0}^{+\infty}\frac{1}{z!}\cosh^m(\nu_a z)+\left(\frac{k}{(k/2)!}+\mathbb{P}\{Z\ge k/2\}\right)\exp(\nu_a mk)\ .\end{aligned}$$
Now we take care separately of these last two terms. First note that
$$\frac{k}{(k/2)!}+\mathbb{P}\{Z\ge k/2\} \le \frac{3k}{(k/2)!} \le \exp((k/3)\log k)$$
when k is large enough, and since $\nu_a mk = O(a^2\rho mk) = o(k\log k)$ by our choice of a, we obtain
$$\left(\frac{k}{(k/2)!}+\mathbb{P}\{Z\ge k/2\}\right)\exp(\nu_a mk)\to 0\ .$$

For the other term the situation is slightly more subtle. Let Y be a sum of m independent Rademacher random variables. Using the binomial identity it is easy to prove that

$$\cosh^m(\nu_a z) = \mathbb{E}\exp(\nu_a z Y) ,$$

and thus we have

$$\frac{1}{e}\sum_{z=0}^{+\infty}\frac{1}{z!}\cosh^m(\nu_a z) = \mathbb{E}\left[\exp(\exp(\nu_a Y) - 1\right] .$$

Now thanks to Hoeffding's inequality, we obtain for any $t > 0$,

$$\mathbb{E}\left[\exp(\exp(\nu_a Y) - 1\right] \leq \exp(\exp(\nu_a t) - 1) + \exp(-t^2)\exp(\exp(\nu_a m) - 1) .$$

In particular with $t = m/\log m$, using that $\nu_a m = O(a^2\rho m) = o(\log m)$ by our choice of a, this shows that

$$\mathbb{E}(\exp(\exp(\nu_a Y) - 1) = 1 + o(1).$$

This leads to the following.

Corollary 4 *Consider the class of perfect matchings on the complete bipartite graph. Under either of* (6.2)*, or* (6.3)*,* $R^* \to 1$*.*

It is easy to derive upper bounds for the performance of the localized squared-sum test in this model. All we need to observe is that $|\mathcal{C}| = k!$ and therefore $\log|\mathcal{C}| \sim k\log k$ when $k \to \infty$. Plugging this into Proposition 2, we see that the local squared-sum test is reliable when

$$\rho_{\text{ave}} \geq (2/k)H^{-1}(3k(\log k)/m)$$

Thus ignoring logarithmic factors, the requirement is that $\rho_{\text{ave}}\sqrt{m\min(k,m)}$ be large. Looking at Corollary 4, the complement of (6.2) *or* (6.3) corresponds to $\rho\sqrt{k\max(k,m)} \to \infty$ *and* $\rho m/\log\min(k,m) \to \infty$ (essentially). Ignoring logarithmic factors, the requirement is that $\rho\sqrt{m\min(k,m)}$ be large. Thus the local squared-sum test is near-optimal.

7 The clique number of random geometric graphs

In this section we describe a, perhaps unexpected, application of Theorem 1. We use this theorem to derive a lower bound for the clique number of random geometric graphs on high-dimensional spheres.

To describe the problem, let $p \in (0,1)$ and let $Z_1,\ldots,Z_n$ be independent random vectors, uniformly distributed on the unit sphere $S_{d-1} = \{x \in \mathbb{R}^d : \|x\| = 1\}$. A random geometric graph $G(n,d,p)$ is defined by vertex set $V = \{1,\ldots,n\}$ and vertex i and vertex j are connected by an edge if and only if $(Z_i, Z_j) \geq t_{p,d}$ where the threshold value $t_{p,d}$ is such that

$$\mathbb{P}\{(Z_1, Z_2) \geq t_{p,d}\} = p$$

(i.e., the probability that an edge is present equals p). The *clique number* $\omega(n,d,p)$ is the size of the largest clique of $G(n,d,p)$ (i.e., the largest fully connected subset of vertices). In

Devroye et al. (2011) the behavior of the random variable $\omega(n,d,p)$ is studied for fixed values of p when n is large and $d = d_n$ grows as a function of n. The rate of growth of $\omega(n,d,p)$ is shown to depend in a crucial way of how fast d_n increases with n. Specifically, the following results are established (and hold with probability converging to 1 as $n \to \infty$):

$$
\begin{array}{lcl}
d_n = O(1) & \Longrightarrow & \omega(n,d,p) = \Omega(n) \\
d_n \to \infty & \Longrightarrow & \omega(n,d,p) = o(n) \\
d_n = o(\log n) & \Longrightarrow & \mathbb{E}\omega(n,d,p) = \Omega(n^{1-\epsilon}) \text{ for all } \epsilon > 0 \\
d_n \geq 9\log^2 n & \Longrightarrow & \omega(n,d,p) = O(\log^3 n) \\
d_n/\log^3 n \to \infty & \Longrightarrow & \omega(n,d,p) = (2+o(1))\log_p n
\end{array}
$$

We see that the clique number behaves in drastically different ways between $d_n = o(\log n)$ — when $\omega(n,d,p)$ grows almost linearly — and $d_n \sim \log^2 n$ — when $\omega(n,d,p)$ has a poly-logarithmic growth at most.

The above-mentioned results leave open the question of where exactly the "phase transition" occurs, and whether the upper bound in the regime $d_n \sim \log^2 n$ is sharp. In this section we are able to answer both of these questions. Below we establish a general lower bound for the clique number which implies that, perhaps surprisingly, the phase transition occurs around $\log^2 n$ and that the upper bounds above cannot be improved in an essential way. We show that the median of the clique number $\omega(n,d,p)$ is bounded from below by $\exp(\kappa \log^2 n/d)$ where κ is a positive constant that depends on p only. This implies, for example, that if $d \sim c\log n$ for some $c > 0$, then $\omega(n,d,p)$ grows as a positive power of n. On the other hand, even when $d \sim \log^{2-\epsilon} n$ for any fixed $\epsilon > 0$, then $\omega(n,d,p)$ is much larger than any power of $\log n$. For the sake of simplicity, we only state the result for the case of $p = 1/2$. The argument is identical for other values of p.

Theorem 2 *There exist universal constants $c_1, c_2, c_3, c_4 > 0$ such that for all n, d such that $d \geq c_1 \log(c_2 n)$, the median of the clique number $\omega(n,d,1/2)$ satisfies*

$$
\mathrm{med}(\omega(n,d,1/2)) \geq c_3 \exp\left(\frac{c_4 \log^2(c_2 n)}{d}\right) .
$$

One may take $c_1 = 7/16$, $c_2 = 16\log 2$, $c_3 = 1/16$, and $c_4 = 49/5120$.

Proof. The basic idea of the proof is to define a test that works well whenever the median clique number is small. But then the lower bound of Theorem 1 implies that the clique number cannot be small.

Let $\omega_0 = \mathrm{med}(\omega(n,d,1/2))$ for short. Consider the clique model with $m = d$, all nonzero correlations equal to ρ and $k = 16\omega_0$. For $i = 1,\dots,n$, let $X^{(i)} = (X_{i,1},\dots,X_{i,d}) \in \mathbb{R}^d$, and define the random geometric graph G on the normalized vectors $Z_i = X^{(i)}/\|X^{(i)}\|$, connecting points Z_i and Z_j whenever $(Z_i, Z_j) \geq 0$. The test statistic we consider is the clique number of the resulting graph, denoted by ω. (This test was suggested and analyzed in Devroye et al. (2011). Here we combine their analysis with Theorem 1 to derive a lower bound for the median clique number.)

Under the null hypothesis (when $\rho = 0$), the Z_i's are i.i.d. uniform on the sphere S_{d-1} implying that $G \sim G(n,d,1/2)$ and, consequently, $\omega \sim \omega(n,d,1/2)$. Devroye et al. (2011)

show that, under the alternative hypothesis, with probability at least 7/8, the graph contains a clique of size k whenever

$$\binom{k}{2} < (1/8)e^{d\rho^2/10}. \tag{7.1}$$

When this is the case, the test that accepts the null hypothesis when $\omega < k$ has a probability of type II error bounded by 1/8. To bound the probability of type I error of this test, we first prove that $\mathbb{E}_0\omega < 2\omega_0$ for any d and n sufficiently large. We start with

$$\mathbb{E}_0\omega \geq 2\omega_0 \quad \Leftrightarrow \quad \mathbb{E}_0\omega - \omega_0 \geq \frac{1}{2}\mathbb{E}_0\omega \quad \Rightarrow \quad \frac{1}{2}\mathbb{E}_0\omega \leq \mathbb{E}_0\omega - \omega_0 \leq \sqrt{\mathrm{var}(\omega)},$$

where in the last step we used the well-known fact that the difference between the mean and the median of any random variable is bounded by its standard deviation. Now observe that ω, as a function of the independent random variables $Z_1, \ldots, Z_n$, is a configuration function in the sense of Talagrand (1995) which implies that $\mathrm{var}(\omega) \leq \mathbb{E}_0\omega$ (Boucheron et al., 2004, Corollary 2). We arrive at

$$\mathbb{E}_0\omega \geq 2\omega_0 \quad \Rightarrow \quad \frac{1}{2}\mathbb{E}_0\omega \leq \sqrt{\mathbb{E}_0\omega} \quad \Leftrightarrow \quad \mathbb{E}_0\omega \leq 4.$$

However, it is a simple matter to show that $\mathbb{E}_0\omega > 4$ for all d if n is sufficiently large. (To see this it suffices to show that the probability that 5 random points form a clique is bounded away from zero. This follows from the arguments of Devroye et al. (2011).) We then bound the probability of type I error as follows

$$\mathbb{P}_0\{\omega \geq k\} = \mathbb{P}_0\{\omega \geq 16\omega_0\} \leq \mathbb{P}_0\{\omega \geq 8\mathbb{E}_0\omega\} \leq \frac{1}{8},$$

where we used Markov's inequality in the last line.

Combining the bounds on the probabilities of type I and type II errors, we conclude that $R^* \leq 1/4$. Put it another way,

$$R^* > 1/4 \Longrightarrow \binom{16\omega_0}{2} \geq (1/8)e^{d\rho^2/10}.$$

Now, by Theorem 1, we see that

$$(16\omega_0)^2 < 4(\ln 2)ne^{-16\rho d/7} \Longrightarrow R^* > 1/4.$$

We conclude that, for any $\rho \in (0, 1)$,

$$(16\omega_0)^2 < 4(\ln 2)ne^{-16\rho d/7} \Longrightarrow (16\omega_0)^2 \geq (1/4)e^{d\rho^2/10}.$$

Therefore, if ρ is such that $4(\ln 2)ne^{-16\rho d/7} > (1/4)e^{d\rho^2/10}$, then $(16\omega_0)^2 \geq (1/4)e^{d\rho^2/10}$. Choosing $\rho = (7/(16d))\log((16\log 2)n)$ — which is possible since $d \geq (7/16)\log((16\log 2)n)$ — clearly satisfies the required inequality and this choice gives rise to the announced lower bound. $\square$

8 Discussion

The cornerstone of our analysis is the lower bound stated in Theorem 1. It is powerful enough that we can deduce useful bounds in many different models, which are seen to be optimal up to constant or logarithmic factors. While a considerable effort has been devoted in the related detection-of-means problem for finding the right constants, one wonders if it is possible to obtain results that fine here, at least in some regimes. One possible avenue for that is via the truncated second moment approach, which underlies the lower bounds in Butucea and Ingster (2011); Donoho and Jin (2004); Hall and Jin (2010); Ingster (1999). The computations are rather daunting in the setup of this paper and we decided not to take that route. Note that the second moment approach (without truncation) has limited applicability, though it is a little more useful here than it is in the case where $m = 1$.

More generally, the problem of detecting correlations of arbitrary sign — not just positive correlations like we do here — remains open. Though one can design natural tests akin to our squared-sum and local squared-sum tests for that situation, the challenge is in deriving tight lower bounds. We mention that our approach to obtaining a lower in Section 2 does not apply here, since the representation (2.1) is not valid when the correlations are negative.

Acknowledgements

EAC is partially supported by ONR grant N00014-09-1-0258. GL is supported by the Spanish Ministry of Science and Technology grant MTM2009-09063 and PASCAL2 Network of Excellence under EC grant no. 216886.

References

Addario-Berry, L., N. Broutin, L. Devroye, and G. Lugosi (2010). On combinatorial testing problems. *Ann. Statist. 38*(5), 3063–3092.

Alon, N., M. Krivelevich, and B. Sudakov (1999). Finding a large hidden clique in a random graph. *Random Structures and Algorithms 13*, 457–466.

Arias-Castro, E., S. Bubeck, and G. Lugosi (2012). Detection of correlations. *Ann. Statist.*. To appear.

Arias-Castro, E., E. J. Candès, H. Helgason, and O. Zeitouni (2008). Searching for a trail of evidence in a maze. *Ann. Statist. 36*(4), 1726–1757.

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

Berman, S. M. (1962). Equally correlated random variables. *Sankhyā Ser. A 24*, 155–156.

Bickel, P. J. and E. Levina (2008a). Covariance regularization by thresholding. *Ann. Statist. 36*(6), 2577–2604.

Bickel, P. J. and E. Levina (2008b). Regularized estimation of large covariance matrices. *Ann. Statist. 36*(1), 199–227.

Boucheron, S., G. Lugosi, and O. Bousquet (2004). Concentration inequalities. *Advanced Lectures on Machine Learning*, 208–240.

Butucea, C. and Y. I. Ingster (2011). Detection of a sparse submatrix of a high-dimensional noisy matrix. Available online http://arxiv.org/abs/1109.0898.

Cai, T. T., C.-H. Zhang, and H. H. Zhou (2010). Optimal rates of convergence for covariance matrix estimation. *Ann. Statist. 38*(4), 2118–2144.

Chen, S. X., L.-X. Zhang, and P.-S. Zhong (2010). Tests for high-dimensional covariance matrices. *J. Amer. Statist. Assoc. 105*(490), 810–819.

Dembo, A. and O. Zeitouni (2010). *Large deviations techniques and applications*, Volume 38 of *Stochastic Modelling and Applied Probability*. Berlin: Springer-Verlag. Corrected reprint of the second (1998) edition.

Devroye, L., A. György, G. Lugosi, and F. Udina (2011). High-dimensional random geometric graphs and their clique number. *Electron. J. Probab. 16*, 2481–2508.

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

El Karoui, N. (2008). Operator norm consistent estimation of large-dimensional sparse covariance matrices. *Ann. Statist. 36*(6), 2717–2756.

Hall, P. and J. Jin (2010). Innovated higher criticism for detecting sparse signals in correlated noise. *Ann. Statist. 38*(3), 1686–1732.

Ingster, Y. I. (1999). Minimax detection of a signal for ℓ_n^l balls. *Math. Methods Statist. 7*, 401–428.

Jin, J. (2003). *Detecting and Estimating Sparse Mixtures*. Ph. D. thesis, Stanford University.

Lam, C. and J. Fan (2009). Sparsistency and rates of convergence in large covariance matrix estimation. *Ann. Statist. 37*(6B), 4254–4278.

Ledoit, O. and M. Wolf (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Ann. Statist. 30*(4), 1081–1102.

Meinshausen, N. and P. Bühlmann (2006). High-dimensional graphs and variable selection with the lasso. *Ann. Statist. 34*(3), 1436–1462.

Muirhead, R. J. (1982). *Aspects of multivariate statistical theory*. New York: John Wiley & Sons Inc. Wiley Series in Probability and Mathematical Statistics.

Rajaratnam, B., H. Massam, and C. M. Carvalho (2008). Flexible covariance estimation in graphical Gaussian models. *Ann. Statist. 36*(6), 2818–2849.

Talagrand, M. (1995). Concentration of measure and isoperimetric inequalities in product spaces. *Publications Mathématiques de l'I.H.E.S. 81*, 73–205.

Verzelen, N. and F. Villers (2010). Goodness-of-fit tests for high-dimensional Gaussian linear models. *Ann. Statist. 38*(2), 704–752.

Yuan, M. and Y. Lin (2007). Model selection and estimation in the Gaussian graphical model. *Biometrika 94*(1), 19–35.