Detecting Markov random fields hidden in white noise

ERY ARIAS-CASTRO[1], SÉBASTIEN BUBECK[2], GÁBOR LUGOSI[3] and
NICOLAS VERZELEN[4]

[1]*Department of Mathematics, University of California, San Diego. E-mail: eariasca@math.ucsd.edu*
[2]*Microsoft Research, Redmond, WA 98052. E-mail: sebubeck@microsoft.com*
[3]*Department of Economics and Business, Pompeu Fabra University, Barcelona, Spain, ICREA, Pg. Llus Companys 23, 08010 Barcelona. E-mail: gabor.lugosi@gmail.com*
[4]*INRA, UMR 729 MISTEA, F-34060 Montpellier, France. E-mail: nicolas.verzelen@inra.fr*

Motivated by change point problems in time series and the detection of textured objects in images, we consider the problem of detecting a piece of a Gaussian Markov random field hidden in white Gaussian noise. We derive minimax lower bounds and propose near-optimal tests.

Keywords: Markov random fields; detection; combinatorial testing; minimax test; image analysis

1. Introduction

Anomaly detection is important in a number of applications, including surveillance and environment monitoring systems using sensor networks, object tracking from video or satellite images, and tumor detection in medical imaging. The most common model is that of an object or signal of unusually high amplitude hidden in noise. In other words, one is interested in detecting the presence of an object in which the mean of the signal is different from that of the background. We refer to this as the *detection-of-means* problem. In many situations, anomaly manifests as unusual dependencies in the data. This *detection-of-correlations* problem is the one that we consider in this paper.

1.1. Setting and hypothesis testing problem

It is common to model dependencies by a Gaussian random field $X = (X_i : i \in \mathcal{V})$, where $\mathcal{V} \subset \mathcal{V}_\infty$ is of size $|\mathcal{V}| = n$, while $\mathcal{V}_\infty$ is countably infinite. We focus on the important example of a d-dimensional integer lattice

$$\mathcal{V} = \{1, \ldots, m\}^d \subset \mathcal{V}_\infty = \mathbb{Z}^d. \tag{1}$$

We formalize the task of detection as the following hypothesis testing problem. One observes a realization of $X = (X_i : i \in \mathcal{V})$, where the X_i's are known to be standard normal. Under the null hypothesis $\mathcal{H}_0$, the X_i's are independent. Under the alternative hypothesis $\mathcal{H}_1$, the X_i's are correlated in one of the following ways. Let $\mathcal{C}$ be a class of subsets of $\mathcal{V}$. Each set $S \in \mathcal{C}$

represents a possible anomalous subset of the components of X. Specifically, when $S \in \mathcal{C}$ is the anomalous subset of nodes, each X_i with $i \notin S$ is still independent of all the other variables, while $(X_i : i \in S)$ coincides with $(Y_i : i \in S)$, where $Y = (Y_i : i \in \mathcal{V}_\infty)$ is a stationary Gaussian Markov random field. We emphasize that, in this formulation, the anomalous subset S is only known to belong to $\mathcal{C}$.

We are thus addressing the problem of detecting a region of a Gaussian Markov random field against a background of white noise. This testing problem models important detection problems such as the detection of a piece of a time series in a signal and the detection of a textured object in an image, which we describe below. Before doing that, we further detail the model and set some foundational notation and terminology.

1.2. Tests and minimax risk

We denote the distribution of X under $\mathcal{H}_0$ by $\mathbb{P}_0$. The distribution of the zero-mean stationary Gaussian Markov random field Y is determined by its covariance operator $\mathbf{\Gamma} = (\mathbf{\Gamma}_{i,j} : i, j \in \mathcal{V}_\infty)$ defined by $\mathbf{\Gamma}_{i,j} = \mathbb{E}[Y_i Y_j]$. We denote the distribution of X under $\mathcal{H}_1$ by $\mathbb{P}_{S,\mathbf{\Gamma}}$ when $S \in \mathcal{C}$ is the anomalous set and $\mathbf{\Gamma}$ is the covariance operator of the Gaussian Markov random field Y.

A *test* is a measurable function $f : \mathbb{R}^{\mathcal{V}} \to \{0, 1\}$. When $f(X) = 0$, the test accepts the null hypothesis and it rejects it otherwise. The probability of *type I* error of a test f is $\mathbb{P}_0\{f(X) = 1\}$. When $S \in \mathcal{C}$ is the anomalous set and Y has covariance operator $\mathbf{\Gamma}$, the probability of *type II* error is $\mathbb{P}_{S,\mathbf{\Gamma}}\{f(X) = 0\}$. In this paper, we evaluate tests based on their *worst-case risks*. The risk of a test f corresponding to a covariance operator $\mathbf{\Gamma}$ and class of sets $\mathcal{C}$ is defined as

$$R_{\mathcal{C},\mathbf{\Gamma}}(f) = \mathbb{P}_0\{f(X) = 1\} + \max_{S \in \mathcal{C}} \mathbb{P}_{S,\mathbf{\Gamma}}\{f(X) = 0\}. \tag{2}$$

Defining the risk this way is meaningful when the distribution of Y is known, meaning that $\mathbf{\Gamma}$ is available to the statistician. In this case, the minimax risk is defined as

$$R_{\mathcal{C},\mathbf{\Gamma}}^* = \inf_f R_{\mathcal{C},\mathbf{\Gamma}}(f), \tag{3}$$

where the infimum is over all tests f. When $\mathbf{\Gamma}$ is only known to belong to some class of covariance operators $\mathfrak{G}$, it is more meaningful to define the risk of a test f as

$$R_{\mathcal{C},\mathfrak{G}}(f) = \mathbb{P}_0\{f(X) = 1\} + \max_{\mathbf{\Gamma} \in \mathfrak{G}} \max_{S \in \mathcal{C}} \mathbb{P}_{S,\mathbf{\Gamma}}\{f(X) = 0\}. \tag{4}$$

The corresponding minimax risk is defined as

$$R_{\mathcal{C},\mathfrak{G}}^* = \inf_f R_{\mathcal{C},\mathfrak{G}}(f). \tag{5}$$

In this paper, we consider situations in which the covariance operator $\mathbf{\Gamma}$ is known (i.e., the test f is allowed to be constructed using this information) and other situations when $\mathbf{\Gamma}$ is unknown but it is assumed to belong to a class $\mathfrak{G}$. When $\mathbf{\Gamma}$ is known (resp. unknown), we say that a test f *asymptotically separates the two hypotheses* if $R_{\mathcal{C},\mathbf{\Gamma}}(f) \to 0$ (resp. $R_{\mathcal{C},\mathfrak{G}}(f) \to 0$), and we say

that the hypotheses *merge asymptotically* if $R_{C,\Gamma}^* \to 1$ (resp. $R_{C,\mathfrak{G}}^* \to 1$), as $n = |\mathcal{V}| \to \infty$. We note that, as long as $\Gamma \in \mathfrak{G}$, $R_{C,\Gamma}^* \le R_{C,\mathfrak{G}}^*$, and that $R_{C,\mathfrak{G}}^* \le 1$, since the test $f \equiv 1$ (which always rejects) has risk equal to 1.

At a high-level, our results are as follows. We characterize the minimax testing risk for both known ($R_{C,\Gamma}^*$) and unknown ($R_{C,\mathfrak{G}}^*$) covariances when the anomaly is a Gaussian Markov random field. More precisely, we give conditions on Γ or $\mathfrak{G}$ enforcing the hypotheses to merge asymptotically so that detection problem is nearly impossible. Under nearly matching conditions, we exhibit tests that asymptotically separate the hypotheses. Our general results are illustrated in the following subsections.

1.3. Example: Detecting a piece of time series

As a first example of the general problem described above, consider the case of observing a time series $X_1, \ldots, X_n$. This corresponds to the setting of the lattice (1) in dimension $d = 1$. Under the null hypothesis, the X_i's are i.i.d. standard normal random variables. We assume that the anomaly comes in the form of temporal correlations over an (unknown) interval $S = \{i + 1, \ldots, i + k\}$ of, say, known length $k < n$. Here, $i \in \{0, 1, \ldots, n - k\}$ is thus unknown. Specifically, when S is the anomalous interval, $(X_{i+1}, \ldots, X_{i+k}) \sim (Y_{i+1}, \ldots, Y_{i+k})$, where $(Y_i : i \in \mathbb{Z})$ is an autoregressive process of order h (abbreviated AR_h) with zero mean and unit variance, that is,

$$Y_i = \psi_1 Y_{i-1} + \cdots + \psi_h Y_{i-h} + \sigma Z_i, \qquad \forall i \in \mathbb{Z}, \tag{6}$$

where $(Z_i : i \in \mathbb{Z})$ are i.i.d. standard normal random variables, $\psi_1, \ldots, \psi_h \in \mathbb{R}$ are the coefficients of the process—assumed to be stationary—and $\sigma > 0$ is such that $\mathrm{Var}(Y_i) = 1$ for all i. Note that σ is a function of $\psi_1, \ldots, \psi_h$, so that the model has effectively h parameters. It is well known that the parameters $\psi_1, \ldots, \psi_h$ define a stationary process when the roots of the polynomial $z^p - \sum_{i=1}^{h} \psi_i z^{p-i}$ in the complex plane lie within the open unit circle. See [14] for a standard reference on time series.

In the simplest setting $h = 1$ and the parameter space for ψ is $(-1, 1)$. Then, the hypothesis testing problem is to distinguish

$$\mathcal{H}_0 : X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1),$$

versus

$$\mathcal{H}_1 : \exists i \in \{0, 1, \ldots, n - k\} \qquad \text{such that}$$

$$X_1, \ldots, X_i, X_{i+k+1}, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$$

and $(X_{i+1}, \ldots, X_{i+k})$ is independent of $X_1, \ldots, X_i, X_{i+k+1}, \ldots, X_n$ with

$$X_{i+j+1} - \psi X_{i+j} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1 - \psi^2), \qquad \forall j \in \{1, \ldots, k - 1\}.$$

Typical realizations of the observed vector under the null and alternative hypotheses are illustrated in Figure 1.
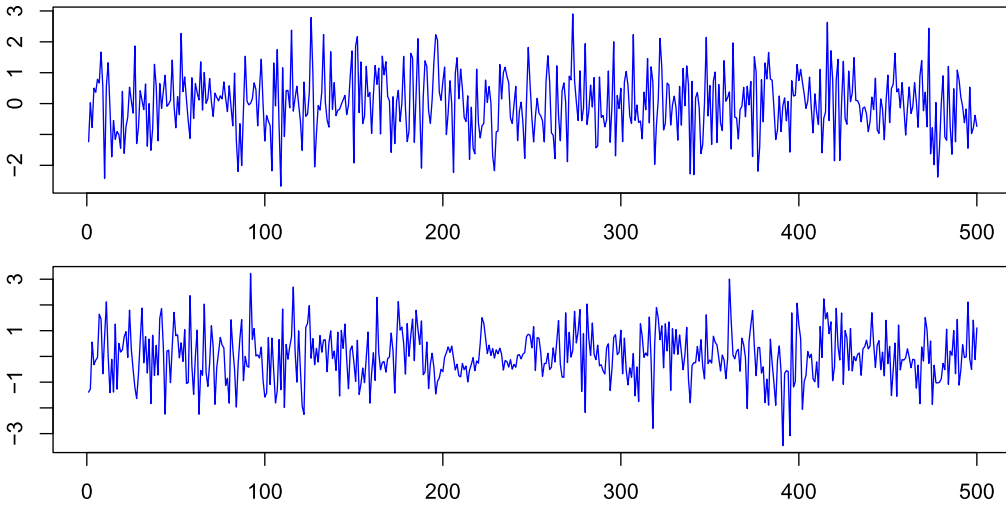
Figure 1. Top: a realization of the observed time series under the null hypothesis (white noise). Bottom: a realization under the alternative with anomalous interval $S = \{201, \ldots, 250\}$, assuming an AR_1 covariance model with parameter $\psi = 0.9$.

Gaussian autoregressive processes and other correlation models are special cases of Gaussian Markov random fields, and therefore this setting is a special case of our general framework, with $\mathcal{C}$ being the class of discrete intervals of length k. In the simplest case, the length of the anomalous interval is known beforehand. In more complex settings, it is unknown, in which case $\mathcal{C}$ may be taken to be the class of all intervals within $\mathcal{V}$ of length at least $k_{\min}$.

This testing problem has been extensively studied in the slightly different context of change-point analysis, where under the null hypothesis $X_1, \ldots, X_n$ are generated from an $AR_h(\psi^0)$ process for some $\psi^0 \in \mathbb{R}^h$, while under the alternative hypothesis there is an $i \in \mathcal{V}$ such that $X_1, \ldots, X_i$ and $X_{i+1}, \ldots, X_n$ are generated from $AR_h(\psi^0)$ and $AR_h(\psi^1)$, with $\psi^0 \neq \psi^1$, respectively. The order h is often given. In fact, instead of assuming autoregressive models, non-parametric models are often favored. See, for example, [19,26,30,32,43,49,50,52] and many other references therein. These papers often suggest maximum likelihood tests whose limiting distributions are studied under the null and (sometimes fixed) alternative hypotheses. For example, in the special case of $h = 1$, such a test would reject $\mathcal{H}_0$ when $|\hat{\psi}|$ is large, where $\hat{\psi}$ is the maximum likelihood estimate for ψ. In particular, from [50], we can speculate that such a test can asymptotically separate the hypotheses in the simplest setting described above when $\psi k^\alpha \to \infty$ for some $\alpha < 1/2$ fixed. See also [32,49] for power analyses against fixed alternatives.

Our general results imply the following in the special case when the anomaly comes in the form of an autoregressive process with unknown parameter $\psi \in \mathbb{R}^h$. We note that the order of the autoregressive model h is allowed to grow with n in this asymptotic result.

Corollary 1. *Assume $n, k \to \infty$, and that $h = o(\sqrt{k/\log(n)} \wedge k^{1/4})$. Denote by $\mathfrak{F}(h, r)$ the class of covariance operators corresponding to AR_h processes with valid parameter $\psi = (\psi_1, \ldots, \psi_h)$*

satisfying $\|\psi\|_2^2 \geq r^2$. *Then* $R_{C,\mathfrak{F}(h,r)}^* \to 1$ *when*

$$r^2 \leq C_1\big(\log(n/k)/k + \sqrt{h\log(n/k)}/k\big). \tag{7}$$

Conversely, if f *denotes the pseudo-likelihood test of Section* 4.2, *then* $R_{C,\mathfrak{F}(h,r)}(f) \to 0$ *when*

$$r^2 \geq C_2\big(\log(n)/k + \sqrt{h\log(n)}/k\big). \tag{8}$$

In both cases, C_1 *and* C_2 *denote numerical constants.*

Remark 1. In the interesting setting where $k = n^\kappa$ for some $\kappa > 0$ fixed, the lower and upper bounds provided by AR match up to a multiplicative constant that depends only on κ.

Despite an extensive literature on the topic, we are not aware of any other minimax optimality result for time series detection.

1.4. Example: Detecting a textured region

In image processing, the detection of textured objects against a textured background is relevant in a number of applications, such as in the detection of local fabric defects in the textile industry by automated visual inspection [40], the detection of a moving object in a textured background [39, 61], the identification of tumors in medical imaging [36,37], the detection of man-made objects in natural scenery [41], the detection of sites of interest in archeology [45] and of weeds in crops [23]. In all these applications, the object is generally small compared to the size of the image.

Common models for texture include Markov random fields [17] and joint distributions over filter banks such as wavelet pyramids [47,51]. We focus here on textures that are generated via Gaussian Markov random fields [16,56]. Our goal is to detect a textured object hidden in white noise. For this discussion, we place ourselves in the lattice setting (1) in dimension $d = 2$. Just like before, under $\mathcal{H}_0$, the $(X_i : i \in \mathcal{V})$ are independent standard normal random variables. Under $\mathcal{H}_1$, when the region $S \subset \mathcal{V}$ is anomalous, the $(X_i : i \notin S)$ are still i.i.d. standard normal, while $(X_i : i \in S) \sim (Y_i : i \in S)$, where $(Y_i : i \in \mathbb{Z}^2)$ is such that for each $i \in \mathbb{Z}^2$, the conditional distribution of Y_i given the rest of the variables $Y^{(-i)} := (Y_j : j \neq i)$ is normal with mean

$$\sum_{(t_1,t_2)\in[-h,h]^2\backslash\{(0,0)\}} \phi_{t_1,t_2} Y_{i+(t_1,t_2)} \tag{9}$$

and variance σ_ϕ^2, where the ϕ_{t_1,t_2}'s are the coefficients of the process and σ_ϕ is such that $\mathrm{Var}(Y_i) = 1$ for all i. The set of valid parameters ϕ is defined in Section 2.1. A simple sufficient condition is $\|\phi\|_1 = \sum_{(t_1,t_2)\in[-h,h]^2\backslash\{(0,0)\}} |\phi_{t_1,t_2}| < 1$. In this model, the *dependency neighborhood* of $i \in \mathbb{Z}^2$ is $i + [-h,h]^2 \cap \mathbb{Z}^2$. One of the simplest cases is when $h = 1$ and $\phi_{t_1,t_2} = \phi$ when $(t_1,t_2) \in \{(\pm1,0),(0,\pm1)\}$ for some $\phi \in (-1/4, 1/4)$, and the anomalous region is a discrete square; see Figure 2 for a realization of the resulting process.
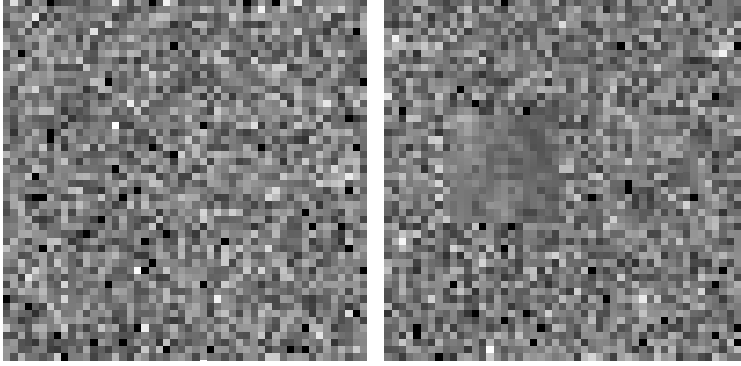
Figure 2. Left: white noise, no anomalous region is present. Right: a squared anomalous region is present. In this example on the 50×50 grid, the anomalous region is a 15×15 square piece from a Gaussian Markov random field with neighborhood radius $h = 1$ and coefficient vector $\phi_{t_1, t_2} = \phi := \frac{1}{4}(1 - 10^{-4})$ when $(t_1, t_2) \in \{(\pm 1, 0), (0, \pm 1)\}$, and zero otherwise.

This is a special case of our setting. While intervals are natural in the case of time series, squares are rather restrictive models of anomalous regions in images. We consider instead the "blob-like" regions (to be defined later) that include convex and star-shaped regions.

A number of publications address the related problems of texture classification [38,56,58] and texture segmentation [25,27,29,35,46]. In fact, this literature is quite extensive. Only very few papers address the corresponding change-point problem [48,55] and we do not know of any theoretical results in this literature. Our general results (in particular, Corollary 4) imply the following.

Corollary 2. *Assume $n, k \to \infty$, and that $h = o(\sqrt{k/\log(n)} \wedge k^{1/5})$. Denote by $\mathfrak{G}(h, r)$ the class of covariance operators corresponding to stationary Gaussian Markov Random Fields with valid parameter (see Section 2.1 for more details) $\phi = (\phi_{i,j})_{(i,j) \in \{-h,...,h\}^2 \setminus \{0\}}$ satisfying $\|\phi\|_2^2 \geq r^2$. Then $R_{\mathcal{C}, \mathfrak{G}(h,r)}^* \to 1$ when*

$$r^2 \leq C_1 \left[\frac{\log(n/k)}{k} + \frac{\sqrt{h^2 \log(n/k)}}{k} \right]. \tag{10}$$

Conversely, if f denotes the pseudo-likelihood test of Section 4.2, then $R_{\mathcal{C}, \mathfrak{G}(h,r)}(f) \to 0$ when

$$r^2 \geq C_2 \left[\frac{\log(n)}{k} + \frac{\sqrt{h^2 \log(n)}}{k} \right]. \tag{11}$$

In both cases, C_1 and C_2 denote positive numerical constants.

Informally, the lower bound on the magnitude of the coefficient vector ϕ, namely r^2, quantifies the extent to which the variables Y_i are explained by the rest of variables $Y^{(-i)}$ as in (9).

Although not in the literature on change-point or object detection, [2] is the only other paper developing theory in a similar context. It considers a spatial model where points $\{x_i, i \in [N]\}$ are sampled uniformly at random in some bounded region and a nearest-neighbor graph is formed. On the resulting graph, variables are observed at the nodes. Under the (simple) null hypothesis, the variables are i.i.d. zero mean normal. Under the (simple) alternative, the variables arise from a Gaussian Markov random field with covariance operator of the form $\Gamma_{i,j} \propto g(\|x_i - x_j\|)$, where g is a known function. The paper analyzes the large-sample behavior of the likelihood ratio test.

1.5. More related work

As we mentioned earlier, the detection-of-means setting is much more prevalent in the literature. When the anomaly has no a priori structure, the problem is that of multiple testing; see, for example, [9,22,34] for papers testing the global null hypothesis. Much closer to what interests us here, the problem of detecting objects with various geometries or combinatorial properties has been extensively analyzed, for example, in some of our earlier work [1,6,7] and elsewhere [20,60]. We only cite a few publications that focus on theory. The applied literature is vast; see [6] for some pointers.

Despite its importance in practice, as illustrated by the examples and references given in Sections 1.3 and 1.4, the detection-of-correlations setting has received comparatively much less attention, at least from theoreticians. Here we find some of our own work [3,4]. In the first of these papers, we consider a sequence $X_1, \ldots, X_n$ of standard normal random variables. Under the null, they are independent. Under the alternative, there is a set S in a class of interest $\mathcal{C}$ where the variables are correlated. We consider the unstructured case where $\mathcal{C}$ is the class of all sets of size k (given) and also various structured cases, and in particular, that of intervals. This would appear to be the same as in the present lattice setting in dimension $d = 1$, but the important difference is that that correlation operator $\mathbf{\Gamma}$ is not constrained, and in particular no Markov random field structure is assumed. The second paper extends the setting to higher dimensions, thus testing whether some coordinates of a high-dimensional Gaussian vector are correlated or not. When the correlation structure in the anomaly is arbitrary, the setting overlaps with that of sparse principal component analysis [10,57]. The problem is also connected to covariance testing in high-dimensions; see, for example, [15]. We refer the reader to the above-mentioned papers for further references.

1.6. Contribution and content

The present paper thus extends previous work on the detection-of-means setting to the detection-of-correlations setting in the (structured) context of detecting signals/objects in time series/images. The paper also extends some of our own work on the detection-of-correlations to Markov random field models, which are typically much more appropriate in the context of detection in signals and images. The theory in the detection-of-correlations setting is more complicated than in the detection-of-means setting, and in particular deriving exact minimax (first-order) results remains an open problem. Compared to our previous work on the detection-of-correlations setting, the Markovian assumption makes the problem significantly more complex as it requires

handling Markov random fields which are conceptually more complex objects. As a result, the proof technique is by-and-large novel, at least in the detection literature.

The rest of the paper is organized as follows. In Section 2, we lay down some foundations on Gaussian Markov Random Fields, and in particular, their covariance operators, and we also derive a general minimax lower bound that is used several times in the paper. In the remainder of the paper, we consider detecting correlations in a finite-dimensional lattice (1), which includes the important special cases of time series and textures in images. We establish lower bounds, both when the covariance matrix is known (Section 3) or unknown (Section 4) and propose test procedures that are shown to achieve the lower bounds up to multiplicative constants. In Section 5, we specialize our general results to specific classes of anomalous regions such as classes of cubes, and more generally, "blobs." In Section 6, we outline possible generalizations and further work. The proofs are gathered in Section 7 and the Supplement.

2. Preliminaries

In this paper, we derive upper and lower bounds for the minimax risk, both when $\boldsymbol{\Gamma}$ is known as in (3) and when it is unknown as in (5), the latter requiring a substantial amount of additional work. For the sake of exposition, we sketch here the general strategy for obtaining minimax lower bounds by adapting the general strategy initiated in [33] to detection-of-correlation problems. This allows us to separate the technique used to derive minimax lower bounds from the technique required to handle Gaussian Markov random fields.

2.1. Some background on Gaussian Markov random fields

We elaborate on the setting described in Sections 1.1 and 1.2. As the process Y is indexed by $\mathbb{Z}^d$, note that all the indices i of ϕ and $\boldsymbol{\Gamma}$ are d-dimensional. Since Y is stationary, all the diagonal elements $\boldsymbol{\Gamma}_{i,i}$ are equal. Given a positive integer h, denote by $\mathbb{N}_h$ the integer lattice $\{-h, \ldots, h\}^d \setminus \{0\}^d$ with $(2h+1)^d - 1$ nodes. For any nonsingular covariance operator $\boldsymbol{\Gamma}$ of a stationary Gaussian Markov random field over $\mathbb{Z}^d$ with unit variance and neighborhood $\mathbb{N}_h$, there exists a unique vector ϕ indexed by the nodes of $\mathbb{N}_h$ satisfying $\phi_i = \phi_{-i}$ such that, for all $i, j \in \mathbb{Z}^d$,

$$\boldsymbol{\Gamma}_{i,j}^{-1}/\boldsymbol{\Gamma}_{i,i}^{-1} = \begin{cases} -\phi_{i-j} & \text{if } 1 \le |i - j|_\infty \le h, \\ 1 & \text{if } i = j, \\ 0 & \text{otherwise,} \end{cases} \tag{12}$$

where $\boldsymbol{\Gamma}^{-1}$ denotes the inverse of the covariance operator $\boldsymbol{\Gamma}$. Consequently, there exists a bijective map from the collection of invertible covariance operators of stationary Gaussian Markov random fields over $\mathbb{Z}^d$ with unit variance and neighborhood $\mathbb{N}_h$ to some subset $\Phi_h \subset \mathbb{R}^{\mathbb{N}_h}$. Given $\phi \in \Phi_h$, $\boldsymbol{\Gamma}(\phi)$ denotes the unique covariance operator satisfying $\boldsymbol{\Gamma}_{i,i} = 1$ and (12). It is well known that Φ_h contains the set of vectors ϕ whose ℓ_1-norm is smaller than one, that is,

$$\left\{ \phi \in \mathbb{R}^{\mathbb{N}_h} : \|\phi\|_1 < 1 \right\} \subset \Phi_h,$$

as the corresponding operator $\mathbf{\Gamma}^{-1}(\phi)$ is diagonally dominant in that case. In fact, the parameter space Φ_h is characterized by the Fast Fourier Transform (FFT) as follows

$$\Phi_h = \left\{ \phi : 1 + \sum_{1 \le |i|_\infty \le h} \phi_i \cos(\langle i, \omega \rangle) > 0, \omega \in (-\pi, \pi]^d \right\},$$

where and $i \in \mathbb{Z}^d$ and $\langle \cdot, \cdot \rangle$ denotes the scalar product in $\mathbb{R}^d$. The interested reader is referred to [28], Section 1.3, or [53], Section 2.6, for further details and discussions. For $\phi \in \Phi_h$, define $\sigma_\phi^2 = 1/\mathbf{\Gamma}_{i,i}^{-1}(\phi)$.

The correlated process $Y = (Y_i : i \in \mathbb{Z}^d)$ is centered Gaussian with covariance operator $\mathbf{\Gamma}(\phi)$ and is such that, for each $i \in \mathbb{Z}^d$, the conditional distribution of Y_i given the rest of the variables $Y^{(-i)}$ is

$$Y_i | Y^{(-i)} \sim \mathcal{N}\left(\sum_{j \in \mathbb{N}_h} \phi_j Y_{i+j}, \sigma_\phi^2 \right). \tag{13}$$

Define the *h-boundary* of S, denoted $\Delta_h(S)$, as the collection of vertices in S whose distance to $\mathbb{Z}^d \setminus S$ is at most h. We also define the *h-interior* S as $S^h = S \setminus \Delta_h(S)$. If $S \subset \mathcal{V}$ is a finite set, we denote by $\mathbf{\Gamma}_S$ the principal submatrix of the covariance operator $\mathbf{\Gamma}$ indexed by S. If $\mathbf{\Gamma}$ is nonsingular, each such submatrix is invertible.

2.2. A general minimax lower bound

As is standard, an upper bound is obtained by exhibiting a test f and then upper-bounding its risk—either (2) or (4) according to whether $\mathbf{\Gamma}$ is known or unknown. In order to derive a lower bound for the minimax risk, we follow the standard argument of choosing a prior distribution on the class of alternatives and then lower-bounding the minimax risk with the resulting *average risk*. When $\mathbf{\Gamma}$ is known, this leads us to select a prior on $\mathcal{C}$, denoted by ν, and consider

$$\bar{R}_{\nu,\mathbf{\Gamma}}(f) = \mathbb{P}_0\big\{ f(X) = 1 \big\} + \sum_{S \in \mathcal{C}} \nu(S) \mathbb{P}_{S,\mathbf{\Gamma}}\big\{ f(X) = 0 \big\} \quad \text{and} \quad \bar{R}_{\nu,\mathbf{\Gamma}}^* = \inf_f \bar{R}_{\nu,\mathbf{\Gamma}}(f). \tag{14}$$

The latter is the *Bayes risk* associated with ν. By placing a prior on the class of alternative distributions, the alternative hypothesis becomes effectively simple (as opposed to composite). The advantage of this is that the optimal test may be determined explicitly. Indeed, the Neyman-Pearson fundamental lemma implies that the likelihood ratio test $f_{\nu,\mathbf{\Gamma}}^*(x) = \mathbb{I}\{L_{\nu,\mathbf{\Gamma}}(x) > 1\}$, with

$$L_{\nu,\mathbf{\Gamma}} = \sum_{S \in \mathcal{C}} \nu(S) \frac{d\mathbb{P}_{S,\mathbf{\Gamma}}}{d\mathbb{P}_0},$$

minimizes the average risk. In most of the paper, ν will be chosen as the uniform distribution on the class $\mathcal{C}$. This is because the sets in $\mathcal{C}$ play almost the same role (although not exactly because of boundary effects).

When $\boldsymbol{\Gamma}$ is only known to belong to some class $\mathfrak{G}$ we also need to choose a prior on $\mathfrak{G}$, which we denote by π, leading to

$$\bar{R}_{\nu,\pi}(f) = \mathbb{P}_0\{f(X) = 1\} + \sum_{S \in \mathcal{C}} \nu(S) \int \mathbb{P}_{S,\boldsymbol{\Gamma}}\{f(X) = 0\}\pi(d\boldsymbol{\Gamma}) \quad \text{and}$$

$$\bar{R}_{\nu,\pi}^* = \inf_f \bar{R}_{\nu,\pi}(f).$$

(15)

In this case, the likelihood ratio test becomes $f_{\nu,\pi}^*(x) = \mathbb{I}\{L_{\nu,\pi}(x) > 1\}$, where

$$L_{\nu,\pi} = \sum_{S \in \mathcal{C}} \nu(S)\frac{d\mathbb{P}_{S,\pi}}{d\mathbb{P}_0}, \qquad \mathbb{P}_{S,\pi} = \int \mathbb{P}_{S,\boldsymbol{\Gamma}}\pi(d\boldsymbol{\Gamma}),$$

minimizes the average risk.

In both cases, we then proceed to bound the second moment of the resulting likelihood ratio under the null. Indeed, in a general setting, if L is the likelihood ratio for $\mathbb{P}_0$ versus $\mathbb{P}_1$ and R denotes its risk, then [44], Problem 3.10

$$R = 1 - \frac{1}{2}\mathbb{E}_0|L(X) - 1| \geq 1 - \frac{1}{2}\sqrt{\mathbb{E}_0[L(X)^2] - 1},$$

(16)

where the inequality follows by the Cauchy–Schwarz inequality.

Remark 2. Working with the minimax risk (as we do here) allows us to bypass making an explicit choice of prior, although one such choice is eventually made when deriving a lower bound. Another advantage is that the minimax risk is monotone with respect to the class $\mathcal{C}$ in the sense that if $\mathcal{C}' \subset \mathcal{C}$, then the minimax risk corresponding to $\mathcal{C}'$ is at most as large as that corresponding to $\mathcal{C}$. This monotonicity does not necessarily hold for the Bayes risk. See [1] for a discussion in the context of the detection-of-means problem.

We now state a general minimax lower bound. (Recall that all the proofs are in Section 7.) Although the result is stated for a class $\mathcal{C}$ of disjoint subsets, using the monotonicity of the minimax risk, the result can be used to derive lower bounds in more general settings. It is particularly useful in the context of detecting blob-like anomalous regions in the lattice. (The same general approach is also fruitful in the detection-of-means setting.) We emphasize that this result is quite straightforward given the work flow outlined above. The technical difficulties will come with its application to the context that interest us here, which will necessitate a good control of (17) below.

Proposition 1. *Let $\{\boldsymbol{\Gamma}(\phi) : \phi \in \Phi\}$ be a class of nonsingular covariance operators and let $\mathcal{C}$ be a class of disjoint subsets of $\mathcal{V}$. Put the uniform prior ν on $\mathcal{C}$ and let π be a prior on Φ. Recall the definition (15) of $\bar{R}_{\nu,\pi}^*$. Then*

$$\bar{R}_{\nu,\pi}^* \geq 1 - \frac{1}{2|\mathcal{C}|}\left(\sum_{S \in \mathcal{C}} V_S\right)^{1/2},$$

where

$$V_S := \mathbb{E}_\pi \left[\left(\frac{\det(\mathbf{\Gamma}_S^{-1}(\phi_1)) \det(\mathbf{\Gamma}_S^{-1}(\phi_2))}{\det(\mathbf{\Gamma}_S^{-1}(\phi_1) + \mathbf{\Gamma}_S^{-1}(\phi_2) - \mathbf{I}_S)} \right)^{1/2} \right], \tag{17}$$

and the expected value is with respect to ϕ_1, ϕ_2 drawn i.i.d. from the distribution π.

3. Known covariance

We start with the case where the covariance operator $\mathbf{\Gamma}$ is known. Although this setting is of less practical importance, as this operator is rarely known in applications, we treat this case first for pedagogical reasons and also to contrast with the much more complex setting where the operator is unknown, treated later on.

3.1. Lower bound

Recall the definition of the minimax risk (3) and the average risk (14). (Henceforth, to lighten the notation, we replace subscripts in $\mathbf{\Gamma}(\phi)$ with subscripts in ϕ.) For any prior ν on $\mathcal{C}$, the minimax risk is at least as large as the ν-average risk, $R^*_{\mathcal{C},\phi} \geq \bar{R}^*_{\nu,\phi}$, and the following corollary of Proposition 1 provides a lower bound on the latter.

Corollary 3. *Let $\mathcal{C}$ be a class of disjoint subsets of $\mathcal{V}$ and fix $\phi \in \Phi_h$ satisfying $\|\phi\|_1 < 1/2$. Then, letting ν denote the uniform prior over $\mathcal{C}$, we have*

$$\bar{R}^*_{\nu,\phi} \geq 1 - \frac{1}{2|\mathcal{C}|} \left[\sum_{S \in \mathcal{C}} \exp\left(\frac{10|S| \|\phi\|_2^2}{1 - 2\|\phi\|_1} \right) \right]^{1/2}. \tag{18}$$

In particular, the corollary implies that, for any fixed $a \in (0, 1)$, $R^*_{\mathcal{C},\phi} \geq 1 - a$ as soon as

$$\frac{\|\phi\|_2^2}{1 - 2\|\phi\|_1} \leq \min_{S \in \mathcal{C}} \frac{\log(4a^2|\mathcal{C}|)}{10|S|}. \tag{19}$$

Furthermore, the hypotheses merge asymptotically (i.e., $R^*_{\mathcal{C},\phi} \to 1$) when

$$\log(|\mathcal{C}|) - \frac{10\|\phi\|_2^2}{1 - 2\|\phi\|_1} \max_{S \in \mathcal{C}} |S| \to \infty. \tag{20}$$

Remark 3. The condition $\|\phi\|_1 < 1/2$ in is technical and likely an artifice of our proof method. This condition arises from the term $\det^{-1/2}(2\mathbf{\Gamma}_S(\phi) - \mathbf{I}_S)$ in V_S in (17). For this determinant to be positive, the smallest eigenvalue of $\mathbf{\Gamma}_S(\phi)$ has to be larger than $1/2$, which in turn is enforced by $\|\phi\|_1 < 1/2$. In order to remove, or at least improve on this constraint, we would need to adopt a more subtle approach than applying the Cauchy–Schwarz inequality in (16). We did not pursue this as typically one is interested in situations where ϕ is small—see, for example, how the result is applied in Section 5.

3.2. Upper bound: The generalized likelihood ratio test

When the covariance operator $\mathbf{\Gamma}(\phi)$ is known, the generalized likelihood ratio test rejects the null hypothesis for large values of

$$\max_{S \in \mathcal{C}} X_S^\top \left(\mathbf{I}_S - \mathbf{\Gamma}_S^{-1}(\phi)\right) X_S.$$

We use instead the statistic

$$U(X) = \max_{S \in \mathcal{C}} \frac{X_S^\top (\mathbf{I}_S - \mathbf{\Gamma}_S^{-1}(\phi)) X_S - \mathrm{Tr}(\mathbf{I}_S - \mathbf{\Gamma}_S^{-1}(\phi))}{\|\mathbf{I}_S - \mathbf{\Gamma}_S^{-1}(\phi)\|_F \sqrt{\log(|\mathcal{C}|)} + \|\mathbf{I}_S - \mathbf{\Gamma}_S^{-1}(\phi)\| \log(|\mathcal{C}|)}, \tag{21}$$

which is based on the centering and normalization of the statistics $X_S^\top (\mathbf{I}_S - \mathbf{\Gamma}_S^{-1}(\phi)) X_S$ where $S \in \mathcal{C}$.

In the following result, we implicitly assume that $|\mathcal{C}| \to \infty$, which is the most interesting case.

Proposition 2. *Assume that* $\phi \in \Phi_h$ *satisfies* $\|\phi\|_1 \leq \eta < 1$ *and that* $|S^h| \geq |S|/2$. *The test* $f(x) = \mathbb{I}\{U(x) > 4\}$ *has risk* $R_{\mathcal{C},\phi}(f) \leq 2/|\mathcal{C}|$ *when*

$$\|\phi\|_2^2 \min_{S \in \mathcal{C}} |S| \geq C_0 \log(|\mathcal{C}|), \tag{22}$$

where $C_0 > 0$ *only depends on the dimension* d *of the lattice and* η.

Comparing with Condition (20), we see that condition (22) matches (up to constants) the minimax lower bound when $\min_S |S|$ is of order $\max_S |S|$, so that (at least when $\|\phi\|_1 < 1/2$) the normalized generalized likelihood ratio test based on (21) is asymptotically minimax up to a multiplicative constant. The ℓ_1-norm $\|\phi\|_1$ arises in the proof of Corollary 3 when bounding the largest eigenvalue of $\mathbf{\Gamma}(\phi)$ (see Lemma 5).

4. Unknown covariance

We now consider the case where the covariance operator $\mathbf{\Gamma}(\phi)$ of the anomalous Gaussian Markov random field is unknown. We therefore start by defining a class of covariance operators via a class of vectors ϕ. Given a positive integer $h > 0$ and some $r > 0$, define

$$\Phi_h(r) := \{\phi \in \Phi_h, \|\phi\|_2 \geq r\}, \tag{23}$$

and let

$$\mathfrak{G}(h, r) := \{\mathbf{\Gamma}(\phi) : \phi \in \Phi_h(r)\}, \tag{24}$$

which is the class of covariance operators corresponding to stationary Gaussian Markov Random Fields with parameter in the class (23).

4.1. Lower bound

The theorem below establishes a lower bound for the risk following the approach outlined in Section 2, which is based on the choice of a suitable prior π on Φ_h, defined as follows. By symmetry of the elements of Φ_h, one can fix a sublattice $\mathbb{N}'_h$ of size $|\mathbb{N}_h|/2$ such that any $\phi \in \Phi_h$ is uniquely defined (via symmetry) by its restriction to $\mathbb{N}'_h$. Choose the distribution π such that $\phi \sim \pi$ is the unique extension to $\mathbb{N}_h$ of the random vector $r|\mathbb{N}_h|^{-1/2}\xi$, where the coordinates of the random vector ξ—indexed by $\mathbb{N}'_h$—are i.i.d. Rademacher random variables (i.e., symmetric ± 1-valued random variables). Note that, if $r|\mathbb{N}_h| < 1$, π is acceptable since it concentrates on the set $\{\phi \in \Phi_h, \|\phi\|_2 = r\} \subset \Phi_h(r)$. Recall the definition of the minimax risk (5) and the average risk (15). As before, for any priors ν on $\mathcal{C}$ and π on $\Phi_h(r)$, the minimax risk is at least as large as the average risk with these priors, $R^*_{\mathcal{C}, \mathfrak{G}(h,r)} \geq \bar{R}^*_{\nu,\pi}$, and the following (much more elaborate) corollary of Proposition 1 provides a lower bound on the latter.

Theorem 1. *There exists a constant $C_0 > 0$ such that the following holds. Let $\mathcal{C}$ be a class of disjoint subsets of $\mathcal{V}$ and let ν denote the uniform prior over $\mathcal{C}$. Let $a \in (0, 1)$ and assume that the neighborhood size $|\mathbb{N}_h|$ satisfies*

$$|\mathbb{N}_h| \leq \min_{S \in \mathcal{C}} \left[\frac{|S|}{\log(a^2|\mathcal{C}|)} \wedge |S|^{2/5} \log^{1/5}(a^2|\mathcal{C}|) \wedge \left(\frac{|S|}{|\Delta_{2h}(S)|} \right)^2 \log^{-1/6}(a^2|\mathcal{C}|) \right]. \tag{25}$$

Then $\bar{R}^*_{\nu,\pi} \geq 1 - a$ as soon as

$$r^2 \max_{S \in \mathcal{C}} |S| \leq C_0 \left[\sqrt{|\mathbb{N}_h| \log(a^2|\mathcal{C}|)} \vee \log(a^2|\mathcal{C}|) \right]. \tag{26}$$

This bound is our main impossibility result. Its proof relies on a number of auxiliary results for Gaussian Markov Random Fields (Section 7.3) that may be useful for other problems of estimating Gaussian Markov Random Fields. Notice that the second term in (26) is what appears in (19), which we saw arises in the case where the covariance is known. In light of this fact, we may interpret the first term in (26) as the 'price to pay' for adapting to an unknown covariance operator in the class of covariance operators of Gaussian Markov random fields with dependency radius h.

4.2. Upper bound: A Fisher-type test

We introduce a test whose performance essentially matches the minimax lower bound of Theorem 1. Comparatively, the construction and analysis of this test is much more involved than that of the generalized likelihood ratio test of Section 3.2.

Let $F_i = (X_{i+v} : 1 \leq |v|_\infty \leq h)$, seen as a vector, and let $\mathbf{F}_{S,h}$ be the matrix with row vectors $F_i, i \in S^h$. Also, let $X_{S,h} = (X_i : i \in S^h)$. Under the null hypothesis, each variable X_i is independent of F_i, although X_i is correlated with some $(F_j, j \neq i)$. Under the alternative hypothesis, there exists a subset S and a vector $\phi \in \Phi_h$ such that

$$X_{S,h} = \mathbf{F}_{S,h}\phi + \varepsilon_{S,h}, \tag{27}$$

where each component ε_i of $\varepsilon_{S,h}$ is independent of the corresponding vector F_i, but the ε_i's are not necessarily independent. Equation (27) is the so-called conditional autoregressive (CAR) representation of a Gaussian Markov random field [28]. For Gaussian Markov random fields, the celebrated pseudo-likelihood method [11] amounts to estimating ϕ by taking least-squares in (27).

Returning to our testing problem, observe that the null hypothesis is true if and only if all the parameters of the conditional expectation of $X_{S,h}$ given $\mathbf{F}_{S,h}$ are zero. In analogy with the analysis-of-variance approach for testing whether the coefficients of a linear regression model are all zero, we consider a Fisher-type statistic

$$T^* = \max_{S \in \mathcal{C}} T_S, \qquad T_S := \frac{|S^h| \, \|\mathbf{\Pi}_{S,h} X_{S,h}\|_2^2}{\|X_{S,h} - \mathbf{\Pi}_{S,h} X_{S,h}\|_2^2}, \tag{28}$$

where $\mathbf{\Pi}_{S,h} := \mathbf{F}_{S,h} (\mathbf{F}_{S,h}^\top \mathbf{F}_{S,h})^{-1} \mathbf{F}_{S,h}^\top$ is the orthogonal projection onto the column space of $\mathbf{F}_{S,h}$. Since in the linear model (27) the response vector $X_{S,h}$ is not independent of the design matrix $\mathbf{F}_{S,h}$, the statistic T_S does not follow an F-distribution. Nevertheless, we are able to control the deviations of T^*, both under null and alternative hypotheses, leading to the following performance bound. Recall the definition (4).

Theorem 2. *There exist four positive constants C_1, C_2, C_3, C_4 depending only on d such that the following holds. Assume that*

$$|\mathbb{N}_h|^4 \vee |\mathbb{N}_h|^2 \log(|\mathcal{C}|) \le C_1 \min_{S \in \mathcal{C}} |S^h|. \tag{29}$$

Fix α and β in $(0, 1)$ such that

$$\log\left(\frac{1}{\alpha}\right) \vee \log\left(\frac{1}{\beta}\right) \le C_2 \frac{\min_{S \in \mathcal{C}} |S^h|}{|\mathbb{N}_h|^2 \log(|\mathcal{C}|)}. \tag{30}$$

Then, under the null hypothesis,

$$\mathbb{P}\left\{ T^* \ge |\mathbb{N}_h| + C_3 \left[\sqrt{|\mathbb{N}_h| \left(\log(|\mathcal{C}|) + 1 + \log(\alpha^{-1}) \right)} + \log(|\mathcal{C}|) + \log(\alpha^{-1}) \right] \right\} \le \alpha, \tag{31}$$

while under the alternative,

$$\mathbb{P}\left\{ T^* \ge |\mathbb{N}_h| + C_4 \left[|S^h| \left(\|\phi\|_2^2 \wedge \frac{1}{|\mathbb{N}^h|} \right) - \sqrt{\mathbb{N}_h} (1 + \log^4(\beta^{-1})) \right] \right\} \ge 1 - \beta. \tag{32}$$

In particular, if $\alpha_n, \beta_n \to 0$ are arbitrary positive sequences, then the test f that rejects the null hypothesis if

$$T^* \ge |\mathbb{N}_h| + C_3 \left[\sqrt{|\mathbb{N}_h| \left(\log(|\mathcal{C}|) + 1 + \log(\alpha_n^{-1}) \right)} + \log(|\mathcal{C}|) + \log(\alpha_n^{-1}) \right]$$

satisfies $R_{\mathcal{C},\mathfrak{G}(h,r)}(f) \to 0$ as soon as

$$r^2 > \frac{C_0}{\min_{S \in \mathcal{C}} |S^h|} \left[\sqrt{|\mathbb{N}_h| \left(\log(|\mathcal{C}|) + \log\left(\frac{1}{\alpha_n}\right) + \log^8\left(\frac{1}{\beta_n}\right) \right)} \vee \log(|\mathcal{C}|) \vee \log\left(\frac{1}{\alpha_n}\right) \right], \quad (33)$$

where $C_0 > 0$ depends only on d.

Comparing with the minimax lower bound established in Theorem 1, we see that this test is nearly optimal with respect to h, the size of the collection $|\mathcal{C}|$, and the size $|S|$ of the anomalous region (under the alternative).

5. Examples: Cubes and blobs

In this section, we specialize our general results proved in the previous subsections to classes of cubes, and more generally, blobs.

5.1. Cubes

Consider the problem of detecting an anomalous cube-shaped region. Let $\ell \in \{1, \ldots, m\}$ and assume that m is an integer multiple of ℓ (for simplicity). Let $\mathcal{C}$ denote the class of all discrete hypercubes of side length ℓ, that is, sets of the form $S = \prod_{s=1}^{d}\{b_s, \ldots, b_s + \ell - 1\}$, where $b_s \in \{1, \ldots, m + 1 - \ell\}$. Each such hypercube $S \in \mathcal{C}$ contains $|S| = k := \ell^d$ nodes, and the class is of size $|\mathcal{C}| = (m - 1 - \ell)^d \leq n$.

The lower bounds for the risk established in Corollary 3 and Theorem 1 are not directly applicable here since these results require subsets of the class $\mathcal{C}$ to be disjoint. However, they apply to any subclass $\mathcal{C}' \subset \mathcal{C}$ of disjoint subsets and, as mentioned in Section 2, any lower bound on the minimax risk over $\mathcal{C}'$ applies to the minimax risk over $\mathcal{C}$. A natural choice for $\mathcal{C}'$ here is that of all cubes of the form $S = \prod_{s=1}^{d}\{a_s\ell + 1, \ldots, (a_s + 1)\ell\}$, where $a_s \in \{0, \ldots, m/\ell - 1\}$. Note that $|\mathcal{C}'| = (m/\ell)^d = n/k$.

h **bounded**. Consider first the case where the radius h of the neighborhood is bounded. We may apply Corollary 3 to get

$$R_{\mathcal{C},\phi}^* \geq 1 - \frac{k^{1/2}}{2n^{1/2}} \exp\left\{ \frac{5k\|\phi\|_2^2}{1 - 2\|\phi\|_1} \right\}.$$

For a given $r > 0$ satisfying $2|\mathbb{N}_h|r \leq 1$, we can choose a parameter ϕ constant over $\mathbb{N}_h$ such that $\|\phi\|_2 = r$ and $\|\phi\|_1 = r\sqrt{(2h+1)^d - 1}$. Since $R_{\mathcal{C},\mathfrak{G}(h,r)}^* \geq R_{\mathcal{C},\phi}^*$, we thus have $R_{\mathcal{C},\mathfrak{G}(h,r)}^* \to 1$ when $n \to \infty$, if $(k, \phi) = (k(n), \phi(n))$ satisfies $\log(n) \ll k \ll n$ and $r^2 \leq \log(n/k)/(11k)$. Comparing with the performance of the Fisher test of Section 4.2, in this particular case, Condition (29) is met, and letting $\alpha = \alpha(n) \to 0$ and $\beta = \beta(n) \to 0$ slowly, we conclude from (33) that this test (denoted f) has risk $R_{\mathcal{C},\mathfrak{G}(h,r)}(f) \to 0$ when $r^2 \geq C_0 \log(n)/k$ for some constant C_0. Thus, in this setting, the Fisher test, without knowledge of ϕ, achieves the correct detection rate as long as $k \leq n^b$ for some fixed $b < 1$.

h **unbounded**. When *h* is unbounded, we obtain a sharper bound by using Theorem 1 instead of Corollary 3. Specialized to the current setting, we derive the following.

Corollary 4. *There exist two positive constants C_1 and C_2 depending only on d such that the following holds. Assume that the neighborhood size h is small enough that*

$$|\mathbb{N}_h| \leq C_1 \left[\frac{k}{\log^{1 \vee (d/2)}(\frac{n}{k})} \wedge k^{2/5} \log^{1/5}\left(\frac{n}{k}\right) \wedge d^{-\frac{2d}{d+2}} k^{\frac{2}{d+2}} \log^{\frac{d}{3d+6}}\left(\frac{n}{k}\right) \right]. \tag{34}$$

Then the minimax risk tends to one when $n \to \infty$ as soon as $(k, h, r) = (k(n), h(n), r(n))$ satisfies $n/k \to \infty$ and

$$r^2 \leq C_2 \left[\frac{\log(\frac{n}{k})}{k} \vee \frac{\sqrt{|\mathbb{N}_h| \log(\frac{n}{k})}}{k} \right]. \tag{35}$$

Note that, in the case of a square neighborhood, $|\mathbb{N}_h| = (2h + 1)^d - 1$. Comparing with the performance of the Fisher test, in this particular case, Condition (29) is equivalent to $|\mathbb{N}_h| \leq C_0 (k^{1/4} \wedge \sqrt{k/\log(n)})$ for some constant C_0. When k is polynomial in n, this condition is stronger than Condition (34) unless $d \leq 5$. In any case, assuming h is small enough that both (29) and (34) hold, and letting $\alpha = \alpha(n) \to 0$ and $\beta = \beta(n) \to 0$ slowly, we conclude from (33) that the Fisher test has risk $R_{\mathcal{C}, \mathfrak{G}(h, r)}$ tending to zero when

$$r^2 \geq C_0 \left[\frac{\log(n)}{k} \vee \frac{\sqrt{|\mathbb{N}_h| \log(n)}}{k} \right],$$

for some large-enough constant $C_0 > 0$, matching the lower bound (35) up to a multiplicative constant as long as $k \leq n^b$ for some fixed $b < 1$.

In conclusion, whether h is fixed or unbounded but growing slowly enough, the Fisher test achieves a separation parameter matching the lower bound up to a multiplicative constant.

5.2. Blobs

So far, we only considered hypercubes, but our results generalize immediately to much larger classes of blob-like regions. Here, we follow the same strategy used in the detection-of-means setting, for example, in [6,8,31].

Fix two positive integers $\ell_\circ \leq \ell^\circ$ and let $\mathcal{C}$ be a class of subsets S such that there are hypercubes $S_\circ$ and S°, of respective side lengths $\ell_\circ$ and ℓ°, such that $S_\circ \subset S \subset S^\circ$. Letting $\mathcal{C}_\circ$ and $\mathcal{C}^\circ$ denote the classes of hypercubes of side lengths $\ell_\circ$ and ℓ°, respectively, our lower bound for the worst-case risk associated with the class $\mathcal{C}^\circ$ obtained from Corollary 4 applies directly to $\mathcal{C}$—although not completely obvious, this follows from our analysis—while scanning over $\mathcal{C}_\circ$ in the Fisher test yields the performance stated above for the class of cubes. In particular, if $\ell_\circ/\ell^\circ$ remains bounded away from 0, the problem of detecting a region in $\mathcal{C}$ is of difficulty comparable to detecting a hypercube in $\mathcal{C}_\circ$ or $\mathcal{C}^\circ$.

When the size of the anomalous region k is unknown, meaning that the class $\mathcal{C}$ of interest includes regions of different sizes, we can simply scan over dyadic hypercubes as done in the first step of the multiscale method of [8]. This does not change the rate as there are less than $2n$ dyadic hypercubes. See also [6].

We note that when $\ell_\circ/\ell^\circ = o(1)$, scanning over hypercubes may not be very powerful. For example, for "convex" sets, meaning when

$$\mathcal{C} = \big\{ S = K \cap \mathcal{V} : K \subset \mathbb{R}^d \text{ convex}, |K \cap \mathcal{V}| = k \big\},$$

it is more appropriate to scan over ellipsoids due to John's ellipsoid theorem [24], which implies that for each convex set $K \subset \mathbb{R}^d$, there is an ellipsoid $E \subset K$ such that $\mathrm{vol}(E) \geq d^{-d} \mathrm{vol}(K)$. For the case where $d = 2$ and the detection-of-means problem, [31]—expanding on ideas proposed in [8]—scan over parallelograms, which can be done faster than scanning over ellipses.

Finally, we mention that what we said in this section may apply to other types of regular lattices, and also to lattice-like graphs such as typical realizations of a random geometric graph. See [6,60] for detailed treatments in the detection-of-means setting.

6. Discussion

We provided lower bounds and proposed near-optimal procedures for testing for the presence of a piece of a Gaussian Markov random field. These results constitute some of the first mathematical results for the problem of detecting a textured object in a noisy image. We leave open some questions and generalization of interest.

Adaptation to unknown neighborhood. When the size h of the neighborhood is unknown, we may build a test rejecting the null hypothesis if for some positive h, the Fisher-type test T^* of Section 4.2 with radius h and level $\alpha_h = \frac{6\alpha}{h^2 6\pi^2}$ rejects the null hypothesis. The corresponding test would simultaneously achieve, over all h, the detection rate (33) up to a multiplicative term of $(\log(h) \vee 1)$.

More refined results. We leave behind the delicate and interesting problem of finding the exact detection rates, with tight multiplicative constants. This is particularly appealing for simple settings such as finding an interval of an autoregressive process, as described in Section 1.3. Our proof techniques, despite their complexity, are not sufficiently refined to get such sharp bounds. We already know that, in the detection-of-means setting, bounding the variance of the likelihood ratio does not yield the right constant. The variant which consists of bounding the first two moments of a carefully truncated likelihood ratio, possibly pioneered in [34], is applicable here, but the calculations are quite complicated and we leave them for future research.

Texture over texture. Throughout the paper, we assumed that the background is Gaussian white noise. This is not essential, but makes the narrative and results more accessible. A more general, and also more realistic setting, would be that of detecting a region where the dependency structure is markedly different from the remainder of the image. This setting has been studied in the context of time series, for example, in some of the references given in Section 1.3. However, we are not aware of existing theoretical results in higher-dimensional settings such as in images.

Other dependency structures. We focused on Markov random fields with limited neighborhood range (quantified by h earlier in the paper). This is a natural first step, particularly since these are popular models for time series and textures. However, one could envision studying other

dependency structures, such as short-range dependency, defined in [54] as situations where the covariances are summable in the following sense

$$\sup_{i\in\mathcal{V}_\infty}\sum_{j\in\mathcal{V}_\infty\setminus i}|\boldsymbol{\Gamma}_{i,j}|<\infty.$$

Estimation of the anomalous region. The detection of the presence of an anomalous region is often followed by the estimation of that region. Since we consider a form of max-test (as opposed to sum-tests), the same procedure can be used for estimating the anomalous region. Indeed, suppose one uses a test statistic of the form $\max_{S\in\mathcal{C}} U_S$, typically originating from the generalized likelihood ratio test, or a variant of it. This naturally leads one to using the estimate $\arg\max_{S\in\mathcal{C}} U_S$, which corresponds to maximum likelihood, or a variant of it. A full-fledged performance analysis of this estimator is well beyond the scope of this paper. One may easily check that, if the null hypothesis is rejected by the test, and the "signal" is strong enough for the test to be reliable, then the estimated region and the true region intersect. Going beyond this would require considerably more effort.

7. Proofs

7.1. Proof of Proposition 1

The Bayes risk is achieved by the likelihood ratio test $f^*_{v,\pi}(x)=\mathbb{I}\{L_{v,\pi}(x)>1\}$ where

$$L_{v,\pi}(x)=\frac{1}{|\mathcal{C}|}\sum_{S\in\mathcal{C}}L_S(x),\qquad\text{with }L_S(x)=\int\frac{\mathrm{d}\mathbb{P}_{S,\boldsymbol{\Gamma}(\phi)}(x)}{\mathrm{d}\mathbb{P}_0(x)}\pi(\mathrm{d}\phi).$$

In our Gaussian model,

$$L_S(x)=\mathbb{E}_\pi\left[\exp\left(\frac{1}{2}x_S^\top\left(\mathbf{I}_S-\boldsymbol{\Gamma}_S^{-1}(\phi)\right)x_S-\frac{1}{2}\log\det\left(\boldsymbol{\Gamma}_S(\phi)\right)\right)\right],\tag{36}$$

where the expectation is taken with respect to the random draw of $\phi\sim\pi$. Then, by (16),

$$\bar{R}^*_{v,\pi}=1-\frac{1}{2}\mathbb{E}_0\big|L_{v,\pi}(X)-1\big|\geq1-\frac{1}{2}\sqrt{\mathbb{E}_0\big[L_{v,\pi}(X)^2\big]-1}.\tag{37}$$

(Recall that $\mathbb{E}_0$ stands for expectation with respect to the standard normal random vector X.)

We proceed to bound the second moment of the likelihood ratio under the null hypothesis. Summing over $S,T\in\mathcal{C}$, we have

$$\mathbb{E}_0\big[L_{v,\pi}(X)^2\big]=\frac{1}{|\mathcal{C}|^2}\sum_{S,T\in\mathcal{C}}\mathbb{E}_0\big[L_S(X)L_T(X)\big]$$

$$=\frac{1}{|\mathcal{C}|^2}\sum_{S\neq T}\mathbb{E}_0\big[L_S(X)\big]\mathbb{E}_0\big[L_T(X)\big]+\frac{1}{|\mathcal{C}|^2}\sum_{S\in\mathcal{C}}\mathbb{E}_0\big[L_S^2(X)\big]$$

$$= \frac{|\mathcal{C}| - 1}{|\mathcal{C}|} + \frac{1}{|\mathcal{C}|^2} \sum_{S \in \mathcal{C}} \mathbb{E}_0 \mathbb{E}_\pi \left[\exp\left(X_S^\top \left(\mathbf{I}_S - \frac{1}{2}\mathbf{\Gamma}_S^{-1}(\phi_1) - \frac{1}{2}\mathbf{\Gamma}_S^{-1}(\phi_2) \right) X_S \right. \right.$$

$$\left. \left. - \frac{1}{2} \log \det\left(\mathbf{\Gamma}_S(\phi_1)\right) - \frac{1}{2} \log \det\left(\mathbf{\Gamma}_S(\phi_2)\right) \right) \right]$$

$$\leq 1 + \frac{1}{|\mathcal{C}|^2} \sum_S \mathbb{E}_\pi \left[\exp\left(-\frac{1}{2} \log \det\left(\mathbf{\Gamma}_S^{-1}(\phi_1) + \mathbf{\Gamma}_S^{-1}(\phi_2) - \mathbf{I}_S\right) \right. \right.$$

$$\left. \left. - \frac{1}{2} \log \det\left(\mathbf{\Gamma}_S(\phi_1)\mathbf{\Gamma}_S(\phi_2)\right) \right) \right]$$

$$= 1 + \frac{1}{|\mathcal{C}|^2} \sum_S V_S,$$

where in the second equality we used the fact that $S \neq T$ are disjoint, and therefore $L_S(X)$ and $L_T(X)$ are independent, and in the third we used the fact that $\mathbb{E}_0[L_S(X)] = 1$ for all $S \in \mathcal{C}$.

7.2. Deviation inequalities

Here we collect a few more-or-less standard inequalities that we need in the proofs. We start with the following standard tail bounds for Gaussian quadratic forms. See, for example, Example 2.12 and Exercise 2.9 in [13].

Lemma 1. *Let Z be a standard normal vector in $\mathbb{R}^d$ and let $\mathbf{R}$ be a symmetric $d \times d$ matrix. Then*

$$\mathbb{P}\left\{ Z^\top \mathbf{R} Z - \mathrm{Tr}(\mathbf{R}) \geq 2\|\mathbf{R}\|_F \sqrt{t} + 2\|\mathbf{R}\| t \right\} \leq e^{-t}, \qquad \forall t \geq 0.$$

Furthermore, if the matrix $\mathbf{R}$ is positive semidefinite, then

$$\mathbb{P}\left\{ Z^\top \mathbf{R} Z - \mathrm{Tr}(\mathbf{R}) \leq -2\|\mathbf{R}\|_F \sqrt{t} \right\} \leq e^{-t}, \qquad \forall t \geq 0.$$

Lemma 2. *There exists a positive constant C such that the following holds. For any Gaussian chaos Z up to order 4 and any $t > 0$,*

$$\mathbb{P}\left\{ \left| Z - \mathbb{E}[Z] \right| \geq C \operatorname{Var}^{1/2}(Z) t^2 \right\} \leq e^{-t}.$$

Proof. This deviation inequality is a consequence of the hypercontractivity of Gaussian chaos. More precisely, Theorem 3.2.10 and Corollary 3.2.6 in [21] state that

$$\mathbb{E} \exp\left[\left(\frac{Z - \mathbb{E}[Z]}{C \operatorname{Var}^{1/2}(Z)} \right)^{1/2} \right] \leq 2,$$

where C is a numerical constant. Then, we apply Markov inequality to prove the lemma. □

Lemma 3. *There exists a positive constant C such that the following holds. Let F be a compact set of symmetric $r \times r$ matrices and let $Y \sim \mathcal{N}(0, \mathbf{I}_r)$. For any $t > 0$, the random variable $Z := \sup_{\mathbf{R} \in F} \mathrm{Tr}[\mathbf{R} Y Y^\top]$ satisfies*

$$\mathbb{P}\{Z \geq \mathbb{E}(Z) + t\} \leq \exp\left(-C\left(\frac{t^2}{\mathbb{E}(W)} \wedge \frac{t}{B}\right)\right), \tag{38}$$

where $W := \sup_{\mathbf{R} \in F} \mathrm{Tr}(\mathbf{R} Y Y^\top \mathbf{R})$ and $B := \sup_{\mathbf{R} \in F} \|\mathbf{R}\|$.

A slight variation of this result where Z is replaced by $\sup_{\mathbf{R} \in F} \mathrm{Tr}[\mathbf{R}(Y Y^\top - \mathbf{I}_r)]$ is proved in [59] using the exponential Efron-Stein inequalities of [12]. Their arguments straightforwardly adapt to Lemma 3.

Lemma 4 ([18]). *Let $\mathbf{W}$ be a standard Wishart matrix with parameters (n, d) satisfying $n > d$. Then for any number $0 < x < 1$,*

$$\mathbb{P}\{\lambda^{\max}(\mathbf{W}) \geq n(1 + \sqrt{d/n} + \sqrt{2x/n})^2\} \leq e^{-x},$$
$$\mathbb{P}\{\lambda^{\min}(\mathbf{W}) \leq n(1 - \sqrt{d/n} - \sqrt{2x/n})_+^2\} \leq e^{-x}.$$

7.3. Auxiliary results for Gaussian Markov random fields on the lattice

Here we gather some technical tools and proofs for Gaussian Markov random fields on the lattice. Recall the notation introduced in Section 2.1.

Lemma 5. *For any positive integer h and $\phi \in \Phi_h$ with $\|\phi\|_1 < 1$, we have that if λ is an eigenvalue of the covariance operator $\mathbf{\Gamma}(\phi)$, then*

$$\frac{\sigma_\phi^2}{1 + \|\phi\|_1} \leq \lambda \leq \frac{\sigma_\phi^2}{1 - \|\phi\|_1}.$$

Also, we have

$$\frac{\|\phi\|_2^2}{1 + \|\phi\|_1} \leq \frac{1 - \sigma_\phi^2}{\sigma_\phi^2} \leq \frac{\|\phi\|_2^2}{1 - \|\phi\|_1} \quad \text{and} \quad 1 - \|\phi\|_1 \leq \sigma_\phi^2 \leq 1. \tag{39}$$

Proof. Recall that $\|\cdot\|$ denotes the $\ell^2 \to \ell^2$ operator norm. First, note that by the definition of ϕ, $\sigma_\phi^2 \mathbf{\Gamma}^{-1}(\phi) - \mathbf{I} = (-\phi_{i-j})_{i,j \in \mathbb{Z}^d}$, and therefore

$$\left\|\sigma_\phi^2 \mathbf{\Gamma}^{-1}(\phi) - \mathbf{I}\right\| \leq \|\phi\|_1, \tag{40}$$

where we used the bound $\|\mathbf{A}\| \leq \sup_{i \in \mathbb{Z}^d} \sum_{j \in \mathbb{Z}^d} |\mathbf{A}_{ij}|$. This implies that the largest eigenvalue of $\mathbf{\Gamma}(\phi)$ is bounded by $\sigma_\phi^2/(1 - \|\phi\|_1)$ if $\|\phi\|_1 < 1$ and that the smallest eigenvalue of $\mathbf{\Gamma}(\phi)$ is at

least $\sigma_\phi^2/(1 + \|\phi\|_1)$. Considering the conditional regression of Y_i given Y_{-i} mentioned in (13), that is,

$$Y_i = \sum_{1 \leq |j|_\infty \leq h} \phi_j Y_{i+j} + \varepsilon_i$$

(with ε_i being standard normal independent of the Y_j for $j \neq i$) and taking the variance of both sides, we obtain

$$1 - \sigma_\phi^2 = \mathrm{Var}\left[\sum_{1 < |j|_\infty \leq h} \phi_j Y_{i+j}\right] = \phi^\top \mathbf{\Gamma}(\phi)\phi \leq \|\mathbf{\Gamma}(\phi)\|\|\phi\|_2^2 \leq \frac{\|\phi\|_2^2}{1 - \|\phi\|_1}\sigma_\phi^2,$$

and therefore

$$1 - \sigma_\phi^2 \leq \frac{\|\phi\|_2^2}{1 - \|\phi\|_1}\sigma_\phi^2.$$

Rearranging this inequality and using the fact that $\|\phi\|_2^2 \leq \|\phi\|_1^2 \leq \|\phi\|_1$, we conclude that $\sigma_\phi^2 \geq 1 - \|\phi\|_1$. The remaining bound $\frac{\|\phi\|_2^2}{1+\|\phi\|_1} \leq \frac{1-\sigma_\phi^2}{\sigma_\phi^2}$ is obtained similarly. $\square$

Recall that for any $v \in \mathbb{Z}^d$, γ_v is the correlation between Y_i and Y_{i+v} and is therefore equal to $\mathbf{\Gamma}_{i,i+v}$. This definition does not depend on the node i since $\mathbf{\Gamma}$ is the covariance of a stationary process.

Lemma 6. *For any h and any $\phi \in \Phi_h$, let $Y \sim \mathcal{N}(0, \mathbf{\Gamma}(\phi))$. As long as $\|\phi\|_1 < 1$, the ℓ_2 norm of the correlations satisfies*

$$\sum_{v \neq 0} \gamma_v^2 \leq \frac{\|\phi\|_2^2}{(1 - \|\phi\|_1)^2} + \left(\frac{\|\phi\|_2^2 \sigma_\phi^2}{(1 - \|\phi\|_1)^2}\right)^2. \tag{41}$$

Proof. In order to compute $\|\gamma\|_2^2$, we use the spectral density of Y defined by

$$f(\omega_1, \ldots, \omega_d) = \frac{1}{(2\pi)^d} \sum_{(v_1, \ldots, v_d) \in \mathbb{Z}^d} \gamma_{v_1, \ldots, v_d} \exp\left(\iota \sum_{i=1}^d v_i \omega_i\right), \qquad (\omega_1, \ldots, \omega_d) \in (-\pi, \pi]^d.$$

Following [28], Section 1.3, or [53], Section 2.6.5, we express the spectral density in terms of ϕ and σ_ϕ^2:

$$\frac{1}{f(\omega_1, \ldots, \omega_d)} = \frac{(2\pi)^d}{\sigma_\phi^2}\left[1 - \sum_{v, 1 \leq |v|_\infty \leq h \in \mathbb{Z}^d} \phi_v e^{\iota\langle v, \omega\rangle}\right],$$

where $\langle \cdot, \cdot \rangle$ denotes the scalar product in $\mathbb{R}^d$. As a consequence,

$$\left|f(\omega_1, \ldots, \omega_d)\right| \leq \sigma_\phi^2\left[(2\pi)^d(1 - \|\phi\|_1)\right]^{-1}.$$

Relying on Parseval formula, we conclude

$$
\sum_{v \neq 0} \gamma_v^2 = (2\pi)^d \int_{[-\pi;\pi]^d} \left[f(\omega_1, \ldots, \omega_d) - \frac{1}{(2\pi)^d} \right]^2 d\omega_1 \cdots d\omega_d
$$

$$
\leq \frac{\sigma_\phi^4}{(2\pi)^d (1 - \|\phi\|_1)^2} \int_{[-\pi;\pi]^d} \left| \frac{1}{(2\pi)^d f(\omega_1, \ldots, \omega_d)} - 1 \right|^2 d\omega_1 \cdots d\omega_d
$$

$$
\leq \frac{\sigma_\phi^4}{(2\pi)^d (1 - \|\phi\|_1)^2} \int_{[-\pi;\pi]^d} \left| \frac{1}{\sigma_\phi^2} - 1 - \frac{1}{\sigma_\phi^2} \sum_{v, 1 \leq |v|_\infty \leq h \in \mathbb{Z}^d} \phi_v e^{\iota \langle v, \omega \rangle} \right|^2 d\omega_1 \cdots d\omega_d
$$

$$
\leq \frac{\sigma_\phi^4}{(2\pi)^d (1 - \|\phi\|_1)^2} \left[(2\pi)^d \left(\frac{1}{\sigma_\phi^2} - 1 \right)^2 + \sum_{v, 1 \leq |v|_\infty} \frac{(2\pi)^d \phi_v^2}{\sigma_\phi^4} \right]
$$

$$
\leq \left(\frac{1 - \sigma_\phi^2}{1 - \|\phi\|_1} \right)^2 + \frac{\|\phi\|_2^2}{(1 - \|\phi\|_1)^2}
$$

$$
\leq \left(\frac{\|\phi\|_2^2 \sigma_\phi^2}{(1 - \|\phi\|_1)^2} \right)^2 + \frac{\|\phi\|_2^2}{(1 - \|\phi\|_1)^2},
$$

where we used (39) in the last line. $\qquad \square$

Lemma 7 (Conditional representation). *For any h and any $\phi \in \Phi_h$, let $Y \sim \mathcal{N}(0, \Gamma(\phi))$. Then for any $i \in \mathbb{Z}^d$, the random variable ε_i defined by the conditional regression $Y_i = \sum_{v \in \mathbb{N}_h} \phi_v Y_{i+v} + \varepsilon_i$ satisfies that:*

1. *ε_i is independent of all X_j, $j \neq i$ and $\mathrm{Cov}(\varepsilon_i, X_i) = \mathrm{Var}(\varepsilon_i) = \sigma_\phi^2$.*
2. *For any $i \neq j$, $\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = -\phi_{i-j} \sigma_\phi^2$ if $|i - j|_\infty \leq h$ and 0 otherwise.*

Proof. The first independence property is a classical consequence of the conditional regression representation for Gaussian random vectors, see, for example, [42]. Since $\mathrm{Var}(\varepsilon_i)$ is the conditional variance of Y_i given $Y^{(-i)}$, it equals $[(\Gamma^{-1}(\phi))_{i,i}]^{-1} = \sigma_\phi^2$. Furthermore,

$$
\mathrm{Cov}(\varepsilon_i, Y_i) = \mathrm{Var}(\varepsilon_i) + \sum_{v \in \mathbb{N}_h} \phi_j \mathrm{Cov}(\varepsilon_i, Y_{i+v}) = \mathrm{Var}(\varepsilon_i),
$$

by the independence of ε_i and $Y^{(-i)}$. Finally, consider any $i \neq j$,

$$
\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = \mathrm{Cov}(\varepsilon_i, Y_j) - \sum_{v \in \mathbb{N}_h} \phi_v \mathrm{Cov}(\varepsilon_i, Y_{j+v}),
$$

where all the terms are equal to zero with the possible exception of $v = i - j$. The result follows. $\qquad \square$

The proofs of the two following lemma are postponed to the supplementary material [5].

Lemma 8 (Comparison of $\Gamma^{-1}(\phi)$ and $\Gamma_S^{-1}(\phi)$). *As long as $\|\phi\|_1 < 1$, the following properties hold*:

1. *If $i \in S^h$ or if $j \in S^h$, then $(\Gamma_S^{-1}(\phi))_{i,j} = (\Gamma^{-1}(\phi))_{i,j}$.*
2. *If $i \in S^h$ and $j \in \Delta_h(S)$, then $1 \leq (\Gamma_S^{-1}(\phi))_{j,j} \leq (\Gamma_S^{-1}(\phi))_{i,i}$.*
3. *If $i \in \Delta_h(S)$, then $\sum_{j \in S: j \neq i} (\Gamma_S^{-1}(\phi))_{i,j}^2 \leq \frac{2\|\phi\|_2^2}{(1-\|\phi\|_1)^3}$.*

Lemma 9. *For any $\phi_1, \phi_2 \in \Phi_h$, define*

$$B_{\phi_1,\phi_2} := \left(\frac{\det(\Gamma_S^{-1}(\phi_1)) \det(\Gamma_S^{-1}(\phi_2))}{\det(\Gamma_S^{-1}(\phi_1) + \Gamma_S^{-1}(\phi_2) - I_S)} \right)^{1/2}.$$

(Note that V_S defined in Proposition 1 equals the expected value of B_{ϕ_1,ϕ_2} when ϕ_1 and ϕ_2 are drawn independently from the distribution π.) Assuming that $\|\phi_1\|_1 \vee \|\phi_2\|_1 < 1/5$, we have

$$\log B_{\phi_1,\phi_2} \leq \frac{1}{2}|S|\langle \phi_1, \phi_2 \rangle + 8Q_S,$$

where

$$Q_S := |S| \sum_{s_1,s_2,s_3=1}^{2} \left| \sum_{j,k \in \mathbb{N}_h} \phi_{s_1,j} \phi_{s_2,k} \phi_{s_3,k-j} \right|$$

$$+ 15|S|(\|\phi_1\|_2^3 \vee \|\phi_2\|_2^3) + |\Delta_h(S)|(\|\phi_1\|_2^2 \vee \|\phi_2\|_2^2)$$

$$+ 28|\Delta_{2h}(S)|(|\Delta_{2h}(S)| \vee (|\mathbb{N}_h| + 1))^{1/2}(\|\phi_1\|_2^3 \vee \|\phi_2\|_2^3).$$

7.4. Proof of Corollary 3

As stated in Lemma 5, all eigenvalues of the covariance operator $\Gamma^{-1}(\phi)$ lie in $(1 - \|\phi\|_1, \frac{1+\|\phi\|_1}{1-\|\phi\|_1})$. Since the spectrum of $\Gamma_S^{-1}(\phi)$ lies between the extrema of the spectrum of $\Gamma^{-1}(\phi)$, and using the assumption that $\|\phi\|_1 < 1/2$, this entails

$$\|\Gamma_S(\phi) - I_S\| \leq \max\left[\frac{2\|\phi\|_1}{1+\|\phi\|_1}, \frac{\|\phi\|_1}{1-\|\phi\|_1} \right] < 1. \tag{42}$$

We now apply Proposition 1 with the probability measure π concentrating on ϕ. In this case,

$$V_S = \frac{\det(\Gamma_S^{-1}(\phi))}{\det(2\Gamma_S^{-1}(\phi) - I_S)^{1/2}} = \det\left(I_S - (I_S - \Gamma_S(\phi))^2\right)^{-1/2},$$

and we get

$$\bar{R}_{v,\phi}^* \geq 1 - \frac{1}{2|\mathcal{C}|} \left[\sum_{S \in \mathcal{C}} \det\left(I_S - (I_S - \Gamma_S(\phi))^2\right)^{-1/2} \right]^{1/2}$$

$$\geq 1 - \frac{1}{2|\mathcal{C}|}\left[\sum_{S \in \mathcal{C}} \exp\left(\frac{\|\mathbf{\Gamma}_S(\phi) - \mathbf{I}_S\|_F^2}{2(1 - \|\mathbf{\Gamma}_S(\phi) - \mathbf{I}_S\|)}\right)\right]^{1/2}$$

$$\geq 1 - \frac{1}{2|\mathcal{C}|}\left[\sum_{S \in \mathcal{C}} \exp\left(\frac{\|\mathbf{\Gamma}_S(\phi) - \mathbf{I}_S\|_F^2}{2(1 - 2\|\phi\|_1)}\right)\right]^{1/2},$$

where $\|\cdot\|_F$ denotes the Frobenius norm. The second inequality above is obtained by applying the inequality $1/(1-\lambda) \leq e^{\lambda/(1-\lambda)}$ for $0 \leq \lambda < 1$ to the eigenvalues of $(\mathbf{\Gamma}_S(\phi) - \mathbf{I}_S)^2$, while the third inequality follows from (42) and the fact that $\|\phi\|_1 < 1/2$. It remains to bound $\|\mathbf{\Gamma}_S(\phi) - \mathbf{I}_S\|_F^2$:

$$\|\mathbf{\Gamma}_S(\phi) - \mathbf{I}_S\|_F^2 = \sum_{(i,j \in S), i \neq j} \mathrm{Cor}^2(Y_i, Y_j)$$

$$\leq |S| \sum_{v \neq 0} \gamma_v^2$$

$$\leq 20|S|\|\phi\|_2^2,$$

where we used Lemma 6, $\sigma_\phi^2 \leq 1$, and $\|\phi\|_2 \leq \|\phi\|_1 \leq 1/2$ in the last line.

7.5. Proof of Theorem 1

Recall the definition of the prior π defined just before the statement of the theorem. Taking the numerical constant C_0 in (26) sufficiently small and relying on condition (25), we have $\|\phi\|_1 = r\sqrt{\mathbb{N}_h} < 1/5$. Consequently, the support of π is a subset of the parameter space Φ_h and we are in position to invoke Lemma 9.

Let ϕ_1, ϕ_2 be drawn independently according to the distribution π and denote by ξ_1 and ξ_2 the corresponding random vectors defined on $\mathbb{N}_h'$. By Lemma 9,

$$\log B_{\phi_1,\phi_2} \leq |S|r^2\mathbb{N}_h^{-1}\langle\xi_1,\xi_2\rangle + 8Q_S,$$

where

$$Q_S \leq 23|S|r^3\sqrt{|\mathbb{N}_h|} + |\Delta_h(S)|r^2 + 28|\Delta_{2h}(S)|\left(|\Delta_{2h}(S)| \vee (|\mathbb{N}_h| + 1)\right)^{1/2}r^3.$$

Since $\langle\xi_1,\xi_2\rangle$ is distributed as the sum of $|\mathbb{N}_h|/2$ independent Rademacher random variables, we deduce that

$$V_S \leq \cosh\left(\frac{r^2|S|}{|\mathbb{N}_h|}\right)^{|\mathbb{N}_h|/2} \exp\left(383\left(|S|\sqrt{|\mathbb{N}_h|} \vee |\Delta_{2h}(S)|^{3/2}\right)r^3 + 8|\Delta_h(S)|r^2\right)$$

$$\leq \exp\left(\frac{r^4|S|^2}{4|\mathbb{N}_h|} \wedge \frac{r^2|S|}{2} + 383\left(|S|\sqrt{|\mathbb{N}_h|} \vee |\Delta_{2h}(S)|^{3/2}\right) + 8|\Delta_h(S)|r^2\right),$$

since $\cosh(x) \leq \exp(x) \wedge \exp(x^2/2)$ for any $x > 0$. Combining this bound with Proposition 1, we conclude that the Bayes risk $\bar{R}^*_{\nu,\pi}$ is bounded from below by

$$1 - \frac{1}{2\sqrt{|\mathcal{C}|}} \left[\max_{S \in \mathcal{C}} \exp\left(\frac{|S|^2 r^4}{4|\mathbb{N}_h|} \wedge \frac{|S| r^2}{2} \right. \right.$$
$$\left. \left. + 383\left(|S|\sqrt{|\mathbb{N}_h| + 1} \vee \left| \Delta_{2h}(S) \right|^{3/2} \right) r^3 + 8\left| \Delta_h(S) \right| r^2 \right) \right]^{1/2}. \tag{43}$$

If the numerical constant C_0 in Condition (26) is sufficiently small, then $\frac{|S|^2 r^4}{4|\mathbb{N}_h|} \wedge \frac{|S| r^2}{2} \leq 0.5 \log(a^2 |\mathcal{C}|)$. Also, choosing C_0 small enough in Condition (26), relying on Condition (25) and on $|\mathbb{N}_h| \geq 1$, we also have

$$383\left(|S|\sqrt{|\mathbb{N}_h| + 1} \vee \left| \Delta_{2h}(S) \right|^{3/2} \right) r^3 + 8\left| \Delta_h(S) \right| r^2 \leq 0.5 \log(a^2 |\mathcal{C}|).$$

Thus, we conclude that $\bar{R}^*_{\nu,\pi} \geq 1 - a/2 \geq 1 - a$.

7.6. Proof of Corollary 4

We deduce the result by closely following the proof of Theorem 1. We first prove that $5r\sqrt{|\mathbb{N}_h|} \leq 1$ is satisfied for n large enough. Starting from (35), we have, for n large enough,

$$25r^2|\mathbb{N}_h| \leq 25C_2 \left(\frac{|\mathbb{N}_h| \log(\frac{n}{k})}{k} \vee \frac{|\mathbb{N}_h|^{3/2} \sqrt{\log(\frac{n}{k})}}{k} \right)$$
$$\leq 25C_2 \left(C_1 \vee \frac{|\mathbb{N}_h|^{3/2} \sqrt{\log(\frac{n}{k})}}{k} \right),$$

where we used Condition (34) in the second line. Taking C_1 and C_2 small enough, we only have to bound $|\mathbb{N}_h|^{3/2} \sqrt{\log(\frac{n}{k})}/k$. We distinguish two cases.

- Case 1: $|\mathbb{N}_h| \leq \log(n/k)$. Since $|\mathbb{N}_h| \leq C_1 k / \log(\frac{n}{k})$, it follows that $|\mathbb{N}_h|^{3/2} \sqrt{\log(\frac{n}{k})}/ k \leq C_1$.
- Case 2: $|\mathbb{N}_h| \geq \log(n/k)$. Then the second part of Condition (34) enforces $\log^{4/5}(n/k) \leq C_1 k^{2/5}$. Using again the second part of Condition (34) yields

$$\frac{|\mathbb{N}_h|^{3/2} \sqrt{\log(\frac{n}{k})}}{k} \leq C_1^{3/2} \frac{\log^{4/5}(n/k)}{k^{2/5}} \leq C_1^{5/2}.$$

As $5r\sqrt{|\mathbb{N}_h|} \leq 1$, we can use the same prior π as in the proof of Theorem 1 and arrive at the same lower bound (43) on R^*_π. It remains to prove that this lower bound goes to one, namely that

$$\frac{|S|^2 r^4}{4|\mathbb{N}_h|} \wedge \frac{|S| r^2}{2} + 383\left(|S|\sqrt{|\mathbb{N}_h| + 1} \vee \left| \Delta_{2h}(S) \right|^{3/2} \right) r^3 + 8\left| \Delta_{2h}(S) \right| r^2 - \log(n/k) \to -\infty,$$

where S is a hypercube of size k. Taking the constant C_2 small enough in (35) leads to $\frac{k^2 r^4}{4|\mathbb{N}_h|} \wedge \frac{kr^2}{2} \leq \log(n/k)/2$.

$$kr^3\sqrt{|\mathbb{N}_h|} \leq C_2^{3/2}\left[\frac{\log(n/k)^3|\mathbb{N}_h|}{k} \vee \frac{\log(n/k)^{3/2}|\mathbb{N}_h|^{5/2}}{k}\right]^{1/2} \leq C_2^{3/2}\big(C_1^{1/2} \vee C_1^{5/4}\big)\log(n/k),$$

where we used again the second part of Condition (34). Taking C_1 and C_2 small enough ensures that $383kr^3\sqrt{|\mathbb{N}_h|+1} \leq \log(n/k)/4$. Finally, it suffices to control $|\Delta_{2h}(S)|^{3/2}r^3$ since $|\Delta_{2h}(S)|r^2 \leq |\Delta_{2h}(S)|^{3/2}r^3 \vee 1$. Observe that

$$\big|\Delta_{2h}(S)\big| = \ell^d - (\ell - 4h)^d = \ell^d\big[1 - (1 - 4h/\ell)^d\big] \leq 4\ell^d\, dh/\ell \leq 4d|\mathbb{N}_h|^{1/d}k^{\frac{d-1}{d}}.$$

It then follows from Condition (35) that

$$\big(d|\mathbb{N}_h|^{1/d}k^{\frac{d-1}{d}}\big)^{3/2}r^3 \leq C_2^{3/2}\left[\frac{d^{3/2}|\mathbb{N}_h|^{3/(2d)}}{k^{3/(2d)}}\log^{1/2}\left(\frac{n}{k}\right) \vee \frac{d^{3/2}|\mathbb{N}_h|^{3/(2d)+3/4}}{k^{3/(2d)}\log^{1/4}(\frac{n}{k})}\right]\log\left(\frac{n}{k}\right)$$

$$\leq C_2^{3/2}\left[C_1^{3/(2d)}d^{3/2}\log^{-1/4}\left(\frac{n}{k}\right)C_1^{\frac{6+3d}{4d}}\right]\log\left(\frac{n}{k}\right),$$

where we used again (34) in the second line. Choosing C_1 and C_2 small enough concludes the proof.

Acknowledgements

Supplementary Material

Supplement to "Detecting Markov random fields hidden in white noise" (DOI: 10.3150/17-BEJ973SUPP; .pdf). This supplement contains the remaining proofs.

References

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

[2] Anandkumar, A., Tong, L. and Swami, A. (2009). Detection of Gauss–Markov random fields with nearest-neighbor dependency. *IEEE Trans. Inform. Theory* **55** 816–827.

[3] Arias-Castro, E., Bubeck, S. and Lugosi, G. (2012). Detection of correlations. *Ann. Statist.* **40** 412–435.

[4] Arias-Castro, E., Bubeck, S. and Lugosi, G. (2015). Detecting positive correlations in a multivariate sample. *Bernoulli* **21** 209–241.

[5] Arias-Castro, E., Bubeck, S., Lugosi, G. and Verzelen, N. (2017). Supplement to "Detecting Markov random fields hidden in white noise." DOI:10.3150/17-BEJ973SUPP.

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

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

[8] 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.

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

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

[11] Besag, J.E. (1975). Statistical Analysis of Non-Lattice Data. *The Statistician* **24** 179–195.

[12] Boucheron, S., Bousquet, O., Lugosi, G. and Massart, P. (2005). Moment inequalities for functions of independent random variables. *Ann. Probab.* **33** 514–560.

[13] Boucheron, S., Lugosi, G. and Massart, P. (2013). *Concentration Inequalities*: *A Nonasymptotic Theory of Independence*. Oxford Univ. Press: London.

[14] Brockwell, P.J. and Davis, R.A. (1991). *Time Series*: *Theory and Methods*, 2nd ed. New York: Springer.

[15] Cai, T.T. and Ma, Z. (2013). Optimal hypothesis testing for high dimensional covariance matrices. *Bernoulli* **19(5B)** 2359–2388.

[16] Chellappa, R. and Chatterjee, S. (1985). Classification of textures using Gaussian Markov random fields. *Acoustics*, *Speech and Signal Processing*, *IEEE Transactions on* **33** 959–963.

[17] Cross, G.R. and Jain, A.K. (1983). Markov random field texture models. *Pattern Analysis and Machine Intelligence*, *IEEE Transactions on* **PAMI-5(1)** 25–39.

[18] Davidson, K.R. and Szarek, S.J. (2001). Local operator theory, random matrices and Banach spaces. In *Handbook of the Geometry of Banach Spaces*, *Vol. I* 317–366. Amsterdam: North-Holland.

[19] Davis, R.A., Wei Huang, D. and Yao, Y.-C. (1995). Testing for a change in the parameter values and order of an autoregressive model. *Ann. Statist.* **23** 282–304.

[20] Desolneux, A., Moisan, L. and Morel, J.-M. (2003). Maximal meaningful events and applications to image analysis. *Ann. Statist.* **31** 1822–1851.

[21] de la Peña, V.H. and Giné, E. (1999). From dependence to independence, randomly stopped processes. *U*-statistics and processes. martingales and beyond. In *Decoupling*. New York: Springer.

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

[23] Dryden, I.L., Scarr, M.R. and Taylor, C.C. (2003). Bayesian texture segmentation of weed and crop images using reversible jump Markov chain Monte Carlo methods. *J. Roy. Statist. Soc. Ser. C* **52** 31–50.

[24] Fritz, J. (1948). Extremum problems with inequalities as subsidiary conditions. In *Studies and Essays Presented to R. Courant on His* 60*th Birthday*, *January* 8 187–204. New York, NY: Interscience Publishers, Inc.

[25] Galun, M., Sharon, E., Basri, R. and Brandt, A. (2003). Texture segmentation by multiscale aggregation of filter responses and shape elements. In *Proceedings IEEE International Conference on Computer Vision* 716–723. Nice, France.

[26] Giraitis, L. and Leipus, R. (1992). Testing and estimating in the change-point problem of the spectral function. *Lith. Math. J.* **32** 15–29.

[27] Grigorescu, S.E., Petkov, N. and Kruizinga, P. (2002). Comparison of texture features based on Gabor filters. *IEEE Trans*. *Image Process*. **11** 1160–1167.

[28] Guyon, X. (1995). *Random Fields on a Network*. New York: Springer.

[29] Hofmann, T., Puzicha, J. and Buhmann, J.M. (1998). Unsupervised texture segmentation in a deterministic annealing framework. *IEEE Trans. Pattern Anal. Mach. Intell.* **20** 803–818.

[30] Horváth, L. (1993). Change in autoregressive processes. *Stochastic Process*. *Appl*. **44** 221–242.

[31] Huo, X. and Ni, X. (2009). Detectability of convex-shaped objects in digital images, its fundamental limit and multiscale analysis. *Statist. Sinica* **19** 1439–1462.

[32] Hušková, M., Prášková, Z. and Steinebach, J. (2007). On the detection of changes in autoregressive time series. I. Asymptotics. *J. Statist*. *Plann*. *Inference* **137** 1243–1259.

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

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

[35] Jain, A.K. and Farrokhnia, F. (1991). Unsupervised texture segmentation using Gabor filters. *Pattern Recognit*. **24** 1167–1186.

[36] James, D., Clymer, B.D. and Schmalbrock, P. (2001). Texture detection of simulated microcalcification susceptibility effects in magnetic resonance imaging of breasts. *J. Magn*. *Reson*. *Imaging* **13** 876–881.

[37] Karkanis, S.A., Iakovidis, D.K., Maroulis, D.E., Karras, D.A. and Tzivras, M. (2003). Computer-aided tumor detection in endoscopic video using color wavelet features. *IEEE Trans. Inf. Technol. Biomed.* **7** 141–152.

[38] Kervrann, C. and Heitz, F. (1995). A Markov random field model-based approach to unsupervised texture segmentation using local and global spatial statistics. *IEEE Trans*. *Image Process*. **4** 856–862.

[39] Kim, K., Chalidabhongse, T.H., Harwood, D. and Davis, L. (2005). Real-time foreground-background segmentation using codebook model. *Real-Time Imaging* **11** 172–185. Special Issue on Video Object Processing.

[40] Kumar, A. (2008). Computer-vision-based fabric defect detection: A survey. *IEEE Trans*. *Ind. Electron*. **55** 348–363.

[41] Kumar, S. and Hebert, M. (2003). Man-made structure detection in natural images using a causal multiscale random field. *Proc*. *IEEE Comput*. *Soc*. *Conf*. *Comput*. *Vis*. *Pattern Recognit*. **1** 119.

[42] Lauritzen, S.L. (1996). *Graphical Models*. *Oxford Statistical Science Series* **17**. New York: The Clarendon Press.

[43] Lavielle, M. and Ludeña, C. (2000). The multiple change-points problem for the spectral distribution. *Bernoulli* **6** 845–869.

[44] Lehmann, E.L. and Romano, J.P. (2005). *Testing Statistical Hypotheses*, 3rd ed. *Springer Texts in Statistics*. New York: Springer.

[45] Litton, C.D. and Buck, C.E. (1995). The Bayesian approach to the interpretation of archaeological data. *Archaeometry* **37** 1–24.

[46] Malik, J., Belongie, S., Leung, T. and Shi, J. (2001). Contour and texture analysis for image segmentation. *Int*. *J*. *Comput*. *Vis*. **43** 7–27.

[47] Manjunath, B.S. and Ma, W.Y. (1996). Texture features for browsing and retrieval of image data. *IEEE Trans*. *Pattern Anal*. *Mach*. *Intell*. **18** 837–842.

[48] Palenichka, R.M., Zinterhof, P. and Volgin, M. (2000). Detection of local objects in images with textured background by using multiscale relevance function. In *Mathematical Modeling*, *Estimation*, *and Imaging* (D.C. Wilson, H.D. Tagare, F.L. Bookstein, F.J. Preteux and E.R. Dougherty, eds.) **4121** 158–169. Bellingham: SPIE.

[49] Paparoditis, E. (2009). Testing temporal constancy of the spectral structure of a time series. *Bernoulli* **15** 1190–1221.

[50] Picard, D. (1985). Testing and estimating change-points in time series. *Adv. in Appl. Probab.* **17** 841–867.

[51] Portilla, J. and Simoncelli, E.P. (2000). A parametric texture model based on joint statistics of complex wavelet coefficients. *Int. J. Comput. Vis.* **40** 49–70.

[52] Priestley, M.B. and Subba Rao, T. (1969). A test for non-stationarity of time-series. *J. Roy. Statist. Soc. Ser. B* **31** 140–149.

[53] Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields*: *Theory and Applications*. *Monographs on Statistics and Applied Probability*. **104**. Boca Raton, FL: Chapman & Hall/CRC.

[54] Samorodnitsky, G. (2006). Long range dependence. *Found. Trends Stoch. Syst.* **1** 163–257.

[55] Shahrokni, A., Drummond, T. and Fua, P. (2004). Texture boundary detection for real-time tracking. In *Computer Vision—ECCV* 2004 (T. Pajdla and JiríJ. Matas, eds.). *Lecture Notes in Computer Science* **3022** 566–577. Berlin/Heidelberg: Springer.

[56] Song Chun Zhu, Wu, Y. and Mumford, D. (1998). Filters, random fields and maximum entropy (frame): Towards a unified theory for texture modeling. *Int. J. Comput. Vis.* **27** 107–126.

[57] Tony Cai, T., Ma, Z. and Wu, Y. (2013). Sparse PCA: Optimal rates and adaptive estimation. *Ann. Statist.* **41** 3074–3110.

[58] Varma, M. and Zisserman, A. (2005). A statistical approach to texture classification from single images. *Int. J. Comput. Vis.* **62** 61–81.

[59] Verzelen, N. (2010). Adaptive estimation of stationary Gaussian fields. *Ann. Statist.* **38** 1363–1402.

[60] Walther, G. (2010). Optimal and fast detection of spatial clusters with scan statistics. *Ann. Statist.* **38** 1010–1033.

[61] Yilmaz, A., Javed, O. and Shah, M. (2006). Object tracking: A survey. *Acm Computing Surveys* (*CSUR*) **38** 13.