GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

On minimax detection of localized signals from indirect or correlated data

Dissertation

for the award of the degree
"Doctor rerum naturalium"
of the Georg-August-Universität Göttingen

within the doctoral program
"Mathematical Sciences"
of the Georg-August University School of Science (GAUSS)

submitted by

Markus Pohlmann

from Münster

Göttingen, 2021

Thesis advisory comittee:

Prof. Dr. Axel Munk
Institute for Mathematical Stochastics, University of Göttingen

Prof. Dr. Frank Werner
Institute of Mathematics, University of Würzburg

Prof. Dr. Daniel Rudolf
Faculty of Computer Science and Mathematics, University of Passau

Members of the examination board:

Reviewer
Prof. Dr. Axel Munk
Institute for Mathematical Stochastics, University of Göttingen

Second Reviewer
Prof. Dr. Frank Werner
Institute of Mathematics, University of Würzburg

Further members of the examination board:

Prof. Dr. Daniel Rudolf
Faculty of Computer Science and Mathematics, University of Passau

Prof. Dr. Gerlind Plonka-Hoch
Institute for Numerical and Applied Mathematics, University of Göttingen

Prof. Dr. Stephan Huckemann
Institute of Mathematical Stochastics, University of Göttingen

Dr. Housen Li
Institute of Mathematical Stochastics, University of Göttingen

Date of the oral examination: February 4, 2022.

Preface

This is a cumulative thesis consisting of the following two papers.

(1) Farida Enikeeva, Axel Munk, Markus Pohlmann and Frank Werner. Bump detection in the presence of dependency: does it ease or does it load? *Bernoulli*, 26(4):3280-3310, 2020. [23]

(2) Markus Pohlmann, Frank Werner and Axel Munk. Minimax detection of localized signals in statistical inverse problems. *arXiv preprint, arXiv:2112.05648*. [60]

The overall topic is minimax detection of certain localized signals from data which is either corrupted by non-independent noise or is only an indirect noisy measurent of the signal.

In [23], we suppose that we observe n consecutive samples from a Gaussian process, for example a time series given by an ARMA(p, q) model, that is stationary except for possible *bumps*, i.e. short periods of time of slightly increased or decreased mean. We provide the asymptotic minimax detection boundary for the problem of detecting such a bump, i.e. the minimal required height of a bump, such that its presence can be reliably confirmed through statistical testing, as n becomes large. This corresponds to the problem of detecting rectangular signals from data with non-independent noise.

In [60], we investigate the detection of localized signals, and linear combinations of such signals, from indirect and noisy data, i.e. in the context of statistical inverse problems. We are able to derive upper bounds, lower bounds, and asymptotic results for the corresponding detection boundaries, and provide examples focusing on the detection of signals that are linear combinations of wavelets in typical inverse problems, such as numerical differentiation, deconvolution, and the inversion of the Radon transform.

This thesis is stuctured as follows. We will present a short introduction to the core ideas of minimax signal detection in Chapter 1. Afterwards, in Chapters 2 and 3 we will provide brief summaries of the settings and main results of the two papers mentioned above. Each chapter will also include a discussion. Finally, in Appendices A.1 and A.2 we will present the first paper [23] and the corresponding supplementary material [24] as published, and in Appendix B we will present the second paper as it can be found on `arxiv.org`.

Acknowledgements

First of all, I would like to thank my supervisor Prof. Axel Munk for putting his trust in me, and giving me the chance to work on this project. It is because of Prof. Munk's enthusiasm and his unshakable confidence that I have been able to bring this project to a conclusion.

I am especially grateful for the support of my second supervisor, Prof. Frank Werner, who always had an open ear and mind to listen to my ideas and worries, spent a lot of his time on discussion with me, and stayed calm even when I was panicking, when things did not work out as I imagined. Prof. Werner's mathematical background in inverse problems also heavily influenced and advanced this project.

I am happy that I had the chance to work with Prof. Farida Enikeeva from the University of Poitiers, whose cooperation made the first part of my project – and thereby, my first successful scientific publication – possible.

I would like thank the other members of my examination board, Prof. Daniel Rudolf, who was the third member of my thesis advisory commitee, Prof. Gerlind Plonka-Hoch, Prof. Stephan Huckemann, and Dr. Housen Li.

Furthermore, I acknowledge the financial support of the Research Training Group 2088.

The Institute for Mathematical Stochastics has been a great workplace, and I am grateful for the support from the current and former staff, especially Stephanie Westphal, Pia Weibelzahl, Heiner Keilholz, Anna Sabrina Friedrich, and Christian Böhm, as well as Diana Sieber from the RTG 2088.

Speaking of the IMS, I would like to thank my colleagues, especially Christoph Weitkamp, Marcel Klatt, Thomas Staudt and Florian Heinemann, for countless discussions on math, the universe and everything.

Most importantly, however, I would not be where I am now, were it not for the continuous support of my family. At every step of the way, I was able to count on their understanding and advice. For this, and for everything else that they have done for me, I would like to thank my parents, Heike and Klaus Pohlmann, and my brother Tobias Pohlmann.

Finally, I want to thank my girlfriend Carina Rosenlehner, who has been there for me, even during the most stressful times. I could not have done this without you and I can not thank you enough!

List of symbols

Numbers

$\mathbb{N}, \mathbb{Z}, \mathbb{R}, \mathbb{C}$	Natural numbers (excluding 0), integers, real and complex numbers, resp.
$\Re(z), \Im(z)$	Real part and imaginary part, resp., of a complex number z
$\bar{z}$	Complex conjugate of a complex number z

Matrices

A^{-1}	Inverse of a matrix A
$\overline{A}, A^T, A^H$	Complex conjugate, transpose and conjugate transpose, resp., of a matrix A
id_n	$n \times n$ identity matrix

Functions

$\mathbb{1}_S$	Indicator function of a set S

Spaces

ℓ^2	Space of square-summable sequences
$L^2(\mathcal{X}, \mathcal{Y}, \nu)$	Space of functions $f : \mathcal{X} \to \mathcal{Y}$ that are square-integrable w.r.t. ν
$\|\cdot\|_{\mathcal{B}}$	Norm of a normed space $\mathcal{B}$
$\langle \cdot, \cdot \rangle_{\mathcal{H}}$	Inner product of an inner product space $\mathcal{H}$

Stochastics

$\mathbb{E}(X)$	Expected value of a random variable X
$\mathcal{N}(\mu, \Sigma)$	Normal distribution with mean μ and covariance matrix Σ

Asymptotics

$a_n = O(b_n)$	$\exists n_0 \in \mathbb{N}, C > 0$, s.t. $\|a_n\| \leq C b_n$ for $n \geq n_0$
$a_n = o(b_n)$	$a_n / b_n \to 0$ as $n \to \infty$
$a_n \precsim b_n$	$\exists n_0 \in \mathbb{N}$, s.t. $a_n \leq b_n$ for $n \geq n_0$
$a_n \sim b_n$	$\exists C > 0$, s.t. $\limsup_{n \to \infty} \|a_n / b_n\| \leq C$
$a_n \asymp b_n$	$a_n / b_n \to 1$ as $n \to \infty$

Contents

1 Introduction **1**
1.1 Testing and distinguishability . . . 2
1.2 Aim of the study . . . 3
1.3 Related literature . . . 3

2 Bump detection in the presence of dependency **5**
2.1 Model and problem statement . . . 5
2.2 Literature review . . . 7
2.3 Main results . . . 7
2.4 Discussion . . . 8

3 Minimax detection of localized signals in statistical inverse problems **9**
3.1 Model and problem statement . . . 9
3.2 Literature review . . . 10
3.3 Main results . . . 11
3.4 Discussion . . . 12

Appendix **13**

A.1 Bump detection in the presence of dependency: Does it ease or does it load? **15**
1 Introduction . . . 16
2 Main results . . . 21
3 ARMA processes and finite sample results . . . 25
4 Simulations . . . 26
5 Proofs . . . 28

A.2 Supplementary material to "Bump detection in the presence of dependency: Does it ease or does it load?" **39**
6 Supplementary material and additional proofs . . . 39

B Minimax detection of localized signals in statistical inverse problems **47**
1 Introduction . . . 47
2 Preliminaries . . . 50
3 Results . . . 52
4 Simulation study . . . 61
5 Discussion . . . 63
6 Proofs . . . 64

Bibliography **75**

Curriculum Vitae **79**

Chapter 1

Introduction

Let us assume that we are interested in an unknown mathematical object, which we call f. Certain properties of f might be known (or assumed to be true) a priori. However, aside from that, all information about f, that is available to us, consists of measurements contaminated by random noise. In statistics, a fundamental endeavour is to produce well-founded conclusions (inference) about such an unknown f, based on the available information, which we call *data*.

In this thesis, we will typically assume that f is a function, that is an element of some Hilbert space $\mathcal{X}$ (for example a subspace of $L^2(\mathbb{R})$). We assume that the available data is given in the form of *direct* measurements (for example samples of f with added random noise, see Chapter 2) or *indirect* measurements (for example the image of f under some transformation with added random noise, see Chapter 3). The noise will be assumed to be Gaussian.

In general, we suppose that the data is a random object Y_ε taking values in some measurable space $\mathcal{S}_\varepsilon$, where ε is a parameter which could be understand to quantify the accuracy of the data. It could, for example, refer to the number of available samples of f (the more the better), or the noise level (standard deviation) of the added noise (the smaller the better).

Often, the objective is to *recover* (or *estimate*) the signal from the available data, i.e. to construct an estimator $\hat{f}$ from Y_ε, such that $\hat{f}$ is – in some sense – close to f with some probabilistic guarantee. This is, however, not what this thesis is about, since for the kind of objects we are interested in, there is an even more fundamental question.

We suppose the following: There is some known reference element $f_0 \in \mathcal{X}$ (which is usually assumed to be 0 without loss of generality), and it is known that f either coincides with f_0 or is close to f_0 in the sense that it deviates from f_0 only by some slight contamination, i.e. $f = f_0 + \eta$, for some $\eta \in \mathcal{X}$. We call this contamination, given by η, an *anomaly*. We suppose that it is known a priori, that only anomalies from certain classes, i.e. subsets of $\mathcal{X}$, can occur.

This motivates the following question, which will be at the center of this thesis: Can we reliably decide, based on the data Y_ε, whether an anomaly is in fact present? We may also say that we are interested in the problem of *detecting* an anomaly. This problem will be formalized in terms of a *testing problem* below (for precise definitions and an overview on the concepts of hypothesis testing, we refer to Section 1.1).

Note that, loosely speaking, if the set of possible anomalies is very broad, and the data Y_ε is not very accurate, then the probability that the data seems to indicate the existence of one of those possible anomalies, when in fact there is none, is increased. Thus, we specify a set $\mathcal{F}_\varepsilon$, depending on the parameter ε, that is a collection of – in some sense normalized – elements of $\mathcal{X}$, and we restrict ourselves to testing only for anomalies of the form $\eta = \delta u$, where $u \in \mathcal{F}_\varepsilon$ and δ is a (real or complex) scalar. In other words, we consider the testing problem

$$H_0 : f = f_0 \quad \text{against} \quad H_{1,\varepsilon} : f = f_0 + \delta u \text{ for some } u \in \mathcal{F}_\varepsilon,\ |\delta| > \rho_\varepsilon, \tag{1.1}$$

based on the data Y_ε, where ρ_ε is a positive real number.

The answer to the question whether such an anomaly can be reliably detected, will be given in terms of ρ_ε. If ρ_ε is large enough such that tests with small error probabilities (see Section 1.1) for the testing problem (1.1) exist, then we would conclude that anomalies of the form $\eta = \delta u$ with $u \in \mathcal{F}_\varepsilon$ and $|\delta| > \rho_\varepsilon$ can be reliably detected.

We will from now on assume that $f_0 = 0$, and call f the *signal*.

1.1 Testing and distinguishability

Here, we present an abridged overview of the most important notions of testing and (minimax) distinguishability, which will be sufficient in the context of this thesis. For a thorough discussion we refer to the seminal book by Yuri I. Ingster and Irina Suslina [40].

In the above testing problem (1.1), we wish to test the *hypothesis* H_0 against the *alternative* $H_{1,\varepsilon}$, which means making an educated guess (based on the data Y_ε) about the correctness of the hypothesis when compared to the alternative, while keeping the error of wrongly deciding against H_0 under control. A (non-randomized) test for the testing problem (1.1) is a measurable function of the data Y_ε given by

$$\phi_\varepsilon : S_\varepsilon \to \{0, 1\}.$$

The test ϕ can be understood as a decision rule in the following sense: If $\phi(Y_\varepsilon) = 0$, the hypothesis is *accepted*. If $\phi(Y_\varepsilon) = 1$, the hypothesis is *rejected* in favor of the alternative.

Note that, in general, the distribution of Y_ε depends on f. We denote the distribution of Y_ε by $\mathbb{P}_f$. In particular, if H_0 is true, then Y_ε has the distribution $\mathbb{P}_0$. If H_0 is true, i.e. $f = 0$, but $\phi(Y_\varepsilon) = 1$, we call this a *type I error* (the hypothesis is rejected although it is true). The probability to make a type I error is

$$\alpha_\varepsilon(\phi) := \mathbb{P}_0(\phi(Y_\varepsilon) = 1).$$

On the other hand, the alternative might be true, but $\phi(Y_\varepsilon) = 0$. We call this a *type II error* (the hypothesis is accepted although the alternative is true). Let us, for simplicity, introduce the notation $\mathcal{F}_\varepsilon(\rho_\varepsilon) = \{\delta u : u \in \mathcal{F}_\varepsilon, |\delta| \geq \rho_\varepsilon\}$, i.e. the set of all possible anomalies that make up the alternative $H_{1,\varepsilon}$. Note that the alternative $H_{1,\varepsilon}$ is composite, i.e. does not only consist of only one element, and the probability to make a type II error will in general depend on the specific element $f \in \mathcal{F}_\varepsilon(\rho_\varepsilon)$. We denote the type II error probability, given that a specific $f \in \mathcal{F}_\varepsilon(\rho_\varepsilon)$ is the true signal, by

$$\beta_\varepsilon(\phi, f) := \mathbb{P}_f(\phi(Y_\varepsilon) = 0), \quad f \in \mathcal{F}_\varepsilon(\rho_\varepsilon),$$

For such composite alternatives we consider the worst case error given by the maximum type II error probability over $\mathcal{F}_\varepsilon(\rho_\varepsilon)$ for our analysis. We define

$$\gamma_\varepsilon = \gamma_\varepsilon(\rho_\varepsilon) := \inf_{\phi \in \Phi_\varepsilon} \left[\alpha_\varepsilon(\phi) + \sup_{f \in \mathcal{F}_\varepsilon(\rho_\varepsilon)} \beta_\varepsilon(\phi, f) \right],$$

where Φ_ε is the set of all tests for the testing problem "H_0 against $H_{1,\varepsilon}$". If γ_ε is small, then that means that there is a test for the testing problem "H_0 against $H_{1,\varepsilon}$" with both small type I and type II error. We would then conclude, as discussed above, that anomalies of the form $\eta = \delta u$ with $u \in \mathcal{F}_\varepsilon$ and $|\delta| > \rho_\varepsilon$ can be reliably detected (in the minimax sense).

Asymptotics

In this thesis, we will usually (although not exclusively) consider an asymptotic setting (which, depending on the context, means either that $\varepsilon \to \infty$ or $\varepsilon \to 0$). We suppose that a family $(\mathcal{F}_\varepsilon)_\varepsilon$ of subsets of $\mathcal{X}$ and a family $(\rho_\varepsilon)_\varepsilon$ of positive real numbers are given, and we consider the family of alternatives $(H_{1,\varepsilon})_\varepsilon$, each defined as above.

We say that the hypothesis H_0 is asymptotically *distinguishable* (in the minimax sense) from the family of alternatives $(H_{1,\varepsilon})_\varepsilon$ when, asymptotically, $\gamma_\varepsilon \to 0$. If $\gamma_\varepsilon \to 1$, we say that they are *indistinguishable*.

For a prescribed family $(\mathcal{F}_\varepsilon)_\varepsilon$, we are interested in determining the smallest possible values ρ_ε, such that H_0 and $H_{1,\varepsilon}$ are still asymptotically distinguishable, if possible. Assume that families $(\rho_\varepsilon^U)_\varepsilon$ and $(\rho_\varepsilon^L)_\varepsilon$ exist, that satisfy

$$\gamma(\rho_\varepsilon) \to 0 \quad \text{if} \quad \rho_\varepsilon \gtrsim \rho_\varepsilon^U, \quad \text{and} \quad \gamma(\rho_\varepsilon) \to 1 \quad \text{if} \quad \rho_\varepsilon \lesssim \rho_\varepsilon^L,$$

as $\sigma \to 0$. If, additionally, $\rho_\varepsilon^L \asymp \rho_\varepsilon^U$, we call a family $(\rho_\varepsilon^*)_\varepsilon$ that satisfies $\rho_\varepsilon^L \lesssim \rho_\varepsilon^* \lesssim \rho_\varepsilon^U$, the *(asymptotic) minimax detection boundary* (or *threshold*). We may say that $(\rho_\varepsilon^*)_\varepsilon$ separates *detectable* and *undetectable* signals.

It is, however, not always possible to find such a sharp threshold. If the family $(\rho_\varepsilon^*)_\varepsilon$ only satisfies the weaker conditions

$$\gamma(\rho_\varepsilon) \to 0 \quad \text{if} \quad \rho_\varepsilon/\rho_\varepsilon^* \to \infty, \quad \text{and} \quad \gamma(\rho_\varepsilon) \to 1 \quad \text{if} \quad \rho_\varepsilon/\rho_\varepsilon^* \to 0,$$

we call it the *separation rate* of the family of testing problems "H_0 against $H_{1,\varepsilon}$".

1.2 Aim of the study

This thesis presents the results of two papers, both dealing with very different instances of the testing problem introduced above, and both containing aspects that, to the best of our knowledge, have not been at the center of the investigation before.

First, we consider the problem of detecting anomalies given by bumps (by which we mean rectangular functions) within a signal, from a collection of samples of that signal contaminated by non-independent Gaussian noise. We present the minimax detection boundary and discuss, whether the dependence of the observations makes the detection of bumps easier or harder.

Second, we consider the problem of detecting localized anomalies, given for example by collections of wavelets or linear combinations of wavelets, from indirect data. While the detection of signals from indirect data (i.e. in the context of inverse problems) has been investigated before, to the best of our knowledge, the focus has not been on these kinds of localized signals before. We present non-asymptotic as well as asymptotic results.

1.3 Related literature

We refer to the subsequent chapters 2 and 3, as well as the papers [23] and [60], presented in the Appendix, since those contain their own discussions on relevant literature.

Suffice it to say, that minimax detection of signals from observations contaminated by Gaussian noise has been investigated for some decades. Some of the most influential classical works include the seminal series of articles [38] by Yuri Ingster, the results of Ermakov [26], and the book [40] by Yuri Ingster and Irina Suslina, just to name a few. We also refer to the works of Baraud [5], who focused on non-asymptotic results.

Chapter 2

Bump detection in the presence of dependency

In this chapter, we dicuss the results of the first paper [23]. The paper itself, as well as the supplementary material [24], are presented, as they are published, in Appendix A.

Suppose that a fragment of a Gaussian process, i.e. a vector of n consecutive samples of said process, is observed. It is assumed that this process is stationary except that, within short periods of time, its mean may be slighly increased or decreased. We call such a temporary change in the mean a *bump*. Due to the random nature of the process, it is possible that a bump, if its induced change in the mean (its *height*) is too small or its duration (its *length*) is too short, may go by unnoticed.

We provide the asymptotic minimax detection boundary for such a bump. To be precise, we investigate the minimal height of a bump, that is required for its detection (in the sense of Chapter 1), under the assumption that only bumps of a certain (known) length can occur.

2.1 Model and problem statement

The above problem can be formalized as follows. Suppose that we observe a triangular array of random variables, whose rows are given by random vectors $Y_n = \mu_n + \xi_n$, $n \in \mathbb{N}$ given by

$$Y_{i,n} = \mu_{i,n} + \xi_{i,n}, \quad 1 \leq i \leq n,$$

where the noise vector $\xi_n = (\xi_{1,n}, \ldots, \xi_{n,n})^T$ represents n consecutive samples from a centered Gaussian process, i.e. $\xi_n \sim \mathcal{N}(0, \Sigma_n)$ for a sequence of positive definite covariance matrices $\Sigma_n \in \mathbb{R}^{n \times n}$. The mean vector μ_n is either 0 or describes a bump. To that end, we suppose that μ_n is obtained from equidistant samples of a function $m_n : [0,1] \to \mathbb{R}$, that is either 0 or of the form $m_n = \delta_n \mathbb{1}_{I_n}$, where $(\delta_n)_{n \in \mathbb{N}}$ is a sequence of real numbers and $(I_n)_{n \in \mathbb{N}}$ is a sequence of intervals $I_n \subseteq [0,1]$, i.e.

$$\mu_{i,n} = m_n\left(\frac{i}{n}\right) = \begin{cases} \delta_n, & \text{if } \frac{i}{n} \in I_n, \\ 0, & \text{else,} \end{cases}$$

for $1 \leq i \leq n$. We suppose that the covariance matrices Σ_n, $n \in \mathbb{N}$, are known. In fact, we will consider very specific classes of covariance matrices (see Assumption 2.2 below). In addition, we suppose that the *lengths* of the intervals I_n (denoted by $\lambda_n := |I_n|$) are known as well. The question we aim to answer is: What is the minimal size of δ_n, as n becomes large, such that a bump of height δ_n and length λ_n can be reliably detected. As outlined in Chapter 1, this means that we need to investigate the sequence of testing problems given by

$$H_{0,n} : m_n = 0 \quad \text{against} \quad H_{1,n} : m_n = \delta \mathbb{1}_I, \text{ for some } I \in \mathcal{I}(\lambda_n), |\delta| \geq \Delta_n, \tag{2.1}$$

where $(\Delta_n)_{n\in\mathbb{N}}$ is a sequence of positive real numbers and $\mathcal{I}(\lambda_n)$ is the set of all intervals $I \subseteq [0,1]$ of length λ_n, i.e.

$$\mathcal{I}(\lambda_n) = \{[a,b) : a,b \in [0,1],\, b-a = \lambda_n\}\,.$$

Now, it becomes clear, how this problem relates to the general framework of minimax signal detection outlined in Chapter 1. If we set $\mathcal{F}_n = \{\mathbb{1}_I : I \in \mathcal{I}(\lambda_n)\}$, then testing problem (2.1) becomes

$$H_{0,n} : m_n = 0 \quad \text{against} \quad H_{1,n} : m_n = \delta u,\ \text{for some } u \in \mathcal{F}_n, |\delta| > \Delta_n \tag{2.2}$$

We are interested in finding the asymptotic minimax detection boundary, given by the sequence $(\Delta_n^*)_{n\in\mathbb{N}}$, as defined in Section 1.1.

Assumptions

Concernig the lengths λ_n of the intervals I_n, we suppose the following (cf. Assumption 1 in [23] (Appendix A.1)).

Assumption 2.1. *The sequence $(\lambda_n)_{n\in\mathbb{N}}$ satisfies the following.*

(i) $\frac{n\lambda_n}{\log n} \to \infty$ *as* $n \to \infty$.

(ii) $\lambda_n = o\left(\frac{1}{\log n}\right)$ *as* $n \to \infty$.

The first part of Assumption 2.1 specifies the minimal length of the interval I_n. It makes sure that enough observations lie within the interval I_n. The second part, which is an upper bound for the size of the interval I_n, is a technical requirement. We acknowledge that the necessity of this assumption may indicate the possibility that the proofs (presented in Appendix A.1) are not optimized, since this part of Assumption 2.1 is not needed in related literature (see for example [14]). Note that Assumption 2.1 also implies that $\lambda_n \to 0$. While we do not consider the case that $\liminf_{n\to\infty} \lambda_n > 0$ here, note that this case has been discussed in the context of similar problems before (again, see [14]).

We assume that the noise vector ξ_n is a sample of n consecutive observations from a centered and stationary Gaussian process $Z = (Z_t)_{t\in\mathbb{Z}}$. We denote the autocovariance function of Z by γ, i.e. $\gamma(k) = \mathbb{E}[Z_t Z_{t+k}]$ for $k \in \mathbb{Z}$. If we assume that the sequence $(\gamma(k))_{k\in\mathbb{Z}}$ is square summable, then we can define its spectral density $g \in L^2([-1/2,1/2))$ by

$$g(\nu) = \sum_{k\in\mathbb{Z}} \gamma(k) e^{-2\pi i k\nu}, \quad \nu \in [-1/2,1/2).$$

In other words, Σ_n has entries $(\Sigma_n)_{i,j} = \gamma(|i-j|) = g_{|i-j|}$, where g_k, $k \in \mathbb{Z}$ are the Fourier coefficients of g. We assume the following (cf. Assumption 2 in [23] (Appendix A.1)).

Assumption 2.2. *The process $Z = (Z_t)_{t\in\mathbb{Z}}$ has a spectral density $g \in L^2([-1/2,1/2))$, and the following holds.*

(i) *g is periodic, i.e.* $\lim_{\nu\to 1/2} g(\nu) = g(-1/2)$.

(ii) *g is essentially bounded away from 0, i.e.* $\operatorname{essinf}_{\nu\in[-1/2,1/2)} g(\nu) > 0$.

(iii) *There are constants $C > 0$ and $\kappa > 0$ such that $|g_k| \le C(1+|k|)^{-(1+\kappa)}$.*

In the following we will give a brief overview of related literature, and then present our main results.

2.2 Literature review

The problem at hand is an instance of the minimax signal detection setting described in Chapter 1, and, as such, it is related to the literature discussed in Section 1.3.

However, this problem is also closely related to the problem of changepoint detection and estimation. The literature is vast, and we can not hope to cover all of it here. We refer to the classical works of Brodsky and Darkhovsky [11], Carlstein et al. [12] and Csörgő and Horváth [15], just to name a few. For more recent reviews of methods in changepoint detection and estimation we refer to Aue and Horváth [3] and Yu [73]. We highlight the results on changepoint detection from correlated observations by Keshavarz et al. [44], which have been especially helpful for this study.

The problem of *detecting* a bump has received much less attention. We refer to the results of Jeng et al. [42] and Chan and Walther [14] (in which the noise is assumed to be i.i.d.), Enikeeva et al. [25] (in which the noise is assumed to be independent, but heterogeneous). It has also been discussed in the context of the estimation of piecewise constant functions, see for example Frick et al. [27] and Pein et al. [58].

For a broader overview and discussions of possible applications, we refer to the paper itself [23] (see Appendix A.1).

2.3 Main results

The main result of this paper is the following.

Theorem 2.3 (cf. Theorem 1 of [23] (Appendix A.1))**.** *If Assumptions 2.1 and 2.2 hold, then the asymptotic minimax detection boundary for the testing problem* (2.1) *satisfies*

$$\Delta_n^* \asymp \sqrt{\frac{-2g(0)\log\lambda_n}{n\lambda_n}},$$

as $n \to \infty$.

Theorem 2.3 answers the question in the title of this paper. Since $g(0) = 1$, when Z represents standard Gaussian white noise (i.e. $\Sigma_n = \mathrm{id}_n$ for all n), it follows from Theorem 2.3 that the presence of dependency makes detection of a bump easier, when $g(0) < 1$, and harder, when $g(0) > 1$. We continue by discussing our main example.

Application to ARMA(p, q) processes

Suppose that Z is a stationary ARMA(p, q) time series with $p \geq 0$, $q \geq 0$, i.e. for any $t \in \mathbb{Z}$ it holds that

$$\varphi(B)Z_t = \theta(B)\zeta_t,$$

where $\zeta_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$ for $t \in \mathbb{Z}$, the functions φ and θ are polynomials of degree p and q, repectively, defined by

$$\varphi(z) = 1 + \sum_{k=1}^{p} \varphi_k z^k, \quad \theta(z) = 1 + \sum_{k=1}^{q} \theta_k z^k, \quad z \in \mathbb{C},$$

and B is the so-called backshift (or lag) operator, defined by $BX_t = X_{t-1}$ for a time series $(X_t)_{t\in\mathbb{Z}}$. Additionally, we suppose that φ and θ have no common roots, and that all roots of both φ and θ lie outside of the unit ball. In this case, Theorem 2.3 yields (cf. Theorem 4 of [23] (Appendix A.1)) that the asymptotic detection boundary satisfies

$$\Delta_n^* \asymp \left|\frac{1 + \sum_{k=1}^{q} \theta_k}{1 + \sum_{k=1}^{p} \varphi_k}\right| \sqrt{\frac{-2\log\lambda_n}{n\lambda_n}}.$$

2.4 Discussion

In this paper, we have derived the asymptotic minimax detection boundary for the detection of a bump in an otherwise stationary Gaussian process. We have shown that the bump height that is required for (asymptotically) powerful detection depends on the long-run variance of the process. This means that bumps in processes with long-run variance at most 1 are easier to detect than in uncorrelated data, and bumps in processes with long-run variance greater than 1 are harder to detect.

An open question is whether these results can be extended to alternatives allowing for different bumpt lengths. This should probably be the first step of further research. We also acknowledge that Assumption 2.2 may be restrictive.

Own contribution

After their preceeding paper on bump detection in heterogeneous Gaussian regression, Farida Enikeeva, Axel Munk and Frank Werner came up with the idea to this study. I refined the problem statement and devised the main part of the theory and carried out the corresponding proofs in cooperation with Farida Enikeeva and under the guidance of Frank Werner and Axel Munk. Farida Enikeeva then added the section concerning non-asymptotic results, and Frank Werner conducted simulations to complement the theoretical results. The initial draft of the paper was written by me and then improved and finished jointly by all authors.

Chapter 3

Minimax detection of localized signals in statistical inverse problems

This chapter summarizes the main results of the paper [60] that is presented in its entirety, as published on `arxiv.org`, in Appendix B.

Here, we investigate thresholds for the detection of localized signals, such as wavelets, or linear combinations of such signals, when only the noisy image of the signal under a linear transformation is available.

3.1 Model and problem statement

Let $A : \mathcal{X} \to \mathcal{Y}$ be a known bounded linear operator between (real or complex) separable Hilbert spaces $\mathcal{X}$ and $\mathcal{Y}$. Suppose that we observe the random process Y_σ on $\mathcal{Y}$ given by

$$Y_\sigma = Af + \sigma\xi, \tag{3.1}$$

where $\sigma > 0$ represents the noise level, and ξ is a Gaussian white noise on $\mathcal{Y}$ (for details see below). Model (3.1) has to be understood in the sense that, for any $h \in \mathcal{Y}$,

$$Y_\sigma(h) = \langle Af, h\rangle_{\mathcal{Y}} + \sigma\xi(h).$$

The white noise ξ is a linear mapping that satisfies the following.

(1) If $\mathcal{X}$ and $\mathcal{Y}$ are real Hilbert spaces, we suppose that $\xi(h) \sim \mathcal{N}(0, \|h\|_{\mathcal{Y}}^2)$ and $\mathbb{E}\left(\xi(h)\xi(h')\right) = \langle h, h'\rangle_{\mathcal{Y}}$ for all $h, h' \in \mathcal{Y}$.

(2) If $\mathcal{X}$ and $\mathcal{Y}$ are complex Hilbert spaces, instead we suppose that $\xi(h) \sim \mathcal{CN}(0, 2\|h\|_{\mathcal{Y}}^2)$ and $\mathbb{E}(\xi(h)\overline{\xi(h')}) = 2\langle h, h'\rangle_{\mathcal{Y}}$. Here, $X \sim \mathcal{CN}(0,1)$ means that X is distributed according to the standard complex normal distribution, i.e. $X = X_1 + iX_2$ for some $X_1, X_2 \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1/2)$.

Model (3.1) can be viewed as a generalization of models that are widely used. For example, if only operators $A : L^2([0,1], \mathbb{R}) \to L^2([0,1], \mathbb{R})$ are considered, then (3.1) is equivalent to the process X_σ on $[0,1]$ given by

$$dX_\sigma(t) = (Af)(t) + \sigma dW(t),$$

where W is a Wiener process. In addition, discretized data of the form $X_\sigma^n \in \mathbb{R}^n$, for some $n \in \mathbb{N}$, given by

$$X_{\sigma,i}^n = (Af)\left(\frac{i}{n}\right) + \sigma\zeta_i,$$

where $\zeta_1, \dots, \zeta_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ can be expressed in this model as well. To that end, define $\tilde{A} := E_n \circ A$, where E_n is the evaluation operator $E_n : g \mapsto (g(1/n), \dots, g(n/n))$. Then $\tilde{A} : L^2([0,1], \mathbb{R}) \to \mathbb{R}^n$ is a bounded linear operator, and equipping $\mathbb{R}^n$ with the usual euclidean inner product shows that X_σ^n is equivalent to (3.1) (with A exchanged for $\tilde{A}$). While this looks superficially similar to the data considered in Chapter 2, note that, in this chapter, asymptotic results are given in terms of $\sigma \to 0$ (as opposed to $n \to \infty$).

Note that, for any instance of the model (3.1), knowing the process Y_σ is equivalent to knowing the Gaussian sequence $y_\sigma = (y_{\sigma,i})_{i \in \mathbb{N}}$ given by

$$y_{\sigma,i} := Y_\sigma(e_i) = \langle Af, e_i \rangle_{\mathcal{Y}} + \sigma \xi_i, \quad i \in \mathbb{N},$$

where $\{e_i : i \in \mathbb{N}\}$ is a Hilbert basis $\mathcal{Y}$, and where, consequently, $\xi_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ (in the real case) or $\xi_i \overset{\text{i.i.d.}}{\sim} \mathcal{CN}(0,2)$ (in the complex case). In the paper [60], we will almost exclusively work with the sequence y_σ, since it is usually more convenient.

Aim of this study

We consider the sequence of testing problems as in (1.1), where the noise level σ takes the role of the parameter ε. However, here we will not restrict ourselves to the analysis of one very specific class of signals (such as the bumps in Chapter 2). Instead, we suppose that a collection of functions $(u_k)_{k \in I} \subseteq \mathcal{X}$, where I is a countable index set, as well as a family $(I_\sigma)_{\sigma > 0}$ of finite subsets of I are given. We will first consider the problem of detecting a signal that is assumed to be a multiple of some u_k, $k \in I_\sigma$, based on data given by Y_σ, and then, in a second step, consider the problem of detecting a signal that is assumed to be a linear combination of the u_k, $k \in I_\sigma$. If the collection of functions $(u_k)_{k \in I}$ is – in a sense – well-behaved, we will be able to determine the asymptotic detection boundary (in the first case) and the separation rate (in the second case), as $\sigma \to 0$. The paper [60] will also feature non-asymptotic results, which we will not discuss here, for the sake of brevity.

We will be especially interested in collections $(u_k)_{k \in I}$ that have a particularly useful structure, such as wavelets or other frames of the space $\mathcal{X}$.

3.2 Literature review

As this is clearly an instance of the general detection problem discussed in Chapter 1, we again refer to the literature presented in Section 1.3.

The problem of estimating f when only data as in (3.1) is available, is a heavily investigated field of research. However, since this is not the focus of this thesis, we restrict ourselves to naming only a few influential works. We refer to the books of Engl, Hanke and Neubauer [22] and Hanke [32] for an overview of classical regularization methods, and to Donoho [18], Abramovich and Silverman [1] for approaches involving wavelet decompositions. We also mention the more recent methods described by Ebner et al. [21] and Hubmer and Ramlau [34], which make use of more general frame decompositions. The latter has been especially inspiring for this paper.

The literature on minimax testing in settings involving data as in (3.1) is not as vast. We refer to the works of Ingster, Sapatinas and Suslina [39], Ingster, Laurent and Marteau [37], Laurent, Loubes and Marteau [48], Marteau and Mathé [53] and Autin et al. [4]. We also mention the results of Laurent, Loubes and Marteau [49], who consider a setting with heterogeneous variances, which (as they point out themselves) can be viewed as equivalent to an inverse problem setting.

We refer to the full version of our paper [60] given in Appendix B for more references and discussion of possible applications.

3.3 Main results

Alternatives given by finite collections of functions

Given a collection of functions $(u_k)_{k\in I}$, such that $Au_k \neq 0$ for all $k \in I$, and a family $(I_\sigma)_{\sigma>0}$ of finite subsets of I, we consider the case that the sets $\mathcal{F}_\sigma$ are given by

$$\mathcal{F}_\sigma = \left\{ \|Au_k\|_{\mathcal{Y}}^{-1} u_k : k \in I_\sigma \right\}.$$

Thus, we consider the family of testing problems

$$H_0 : f = 0 \quad \text{against} \quad H_{1,\sigma} : f = \delta u \text{ for some } u \in \mathcal{F}_\sigma,\ |\delta| > \mu_\sigma. \tag{3.2}$$

The main result for this section is the following (cf. Corollary 3.3 of [60]).

Theorem 3.1 (cf. Corollary 3.3 in [60] (Appendix B))**.** *Assume that $N_\sigma = |I_\sigma| \to \infty$, and let*

$$M_\sigma = \sup_{k\in I_\sigma} \#\{k' \in I_\sigma : \Re(\langle Au_k, Au_{k'}\rangle_{\mathcal{Y}}) > 0\},$$

and assume that $M_\sigma N_\sigma^{-\varepsilon_\sigma} \to 0$ for a family $(\varepsilon_\sigma)_{\sigma>0}$ that satifies $\varepsilon_\sigma \to 0$ and $\varepsilon_\sigma\sqrt{\log N_\sigma} \to \infty$ as $\sigma \to 0$. Then the asymptotic detection boundary satisfies

$$\mu_\sigma^* \asymp \sqrt{2\sigma^2 \log N_\sigma}.$$

This means that the detection boundary for alternatives given by $\mathcal{F}_\sigma(\mu_\sigma)$ depends on the inner products of the images of the u_k under A. We discuss a few cases in which Theorem 3.1 is applicable in Section 3.1 of [60].

Alternatives given by the linear span of collections of functions

As discussed above, we will now consider the case that a signal may be a linear combination of the u_k, $k \in I_\sigma$. We start this part by assuming the following (cf. Assumption 3.8 in [60]).

Assumption 3.2. *There is a collection $(v_k)_{k\in I}$ of functions in $\mathcal{Y}$, and a sequence $(\lambda_k)_{k\in I}$ of non-zero complex numbers, such that for any $f \in \mathcal{X}$ it holds that*

$$\langle Af, v_k\rangle_{\mathcal{Y}} = \lambda_k \langle f, u_k\rangle_{\mathcal{X}}.$$

Assumption 3.2 introduces the numbers λ_k, $k \in I$, which are often called quasi-singular values, as they are analogous to the singular values of the SVD. This allows us to define

$$\mathcal{F}_\sigma^L = \left\{ f \in \operatorname{span}\{u_k : k \in I_\sigma\} : \sum_{k\in I_\sigma} |\lambda_k \langle f, u_k\rangle_{\mathcal{X}}|^2 = 1 \right\}.$$

We consider the family of testing problems given by

$$H_0 : f = 0 \quad \text{against} \quad H_{1,\sigma} : f = \delta u \text{ for some } u \in \mathcal{F}_\sigma^L,\ |\delta| > \nu_\sigma. \tag{3.3}$$

Assumption 3.2 is enough to derive some non-asymptotic and asymptotic results. For the sake of brevity, we will omit the discussion of these results here, and instead refer to Section 3.2 of [60].

Instead, we will only present the final result from [60], which gives the separation rate for the testing problem (3.2), when the system of functions $(u_k)_{k\in I}$ behaves especially well. We assume the following (cf. Assumption 3.12 in [60]).

Assumption 3.3. *The collections of functions $(v_k)_{k\in I}$ and $(\tilde{v}_k)_{k\in I}$ defined by $\tilde{v}_k = \lambda_k^{-1} Au_k$ are biorthogonal Riesz sequences.*

Then the following holds.

Theorem 3.4 (cf. Corollary 3.11 and Lemma 3.13 in [60] (Appendix B))**.** *If Assumptions 3.2 and 3.3 hold, then the separation rate of the testing problem* (3.3) *satisfies*

$$\nu_\sigma^* \sim \sigma N_\sigma^{1/4},$$

where $N_\sigma = |I_\sigma|$.

Application to the wavelet-vaguelette decomposition

For specific operators A (such as integration, convolution and the Radon transform), it was shown by David Donoho [18] that, if the collection $(u_k)_{k\in I}$ is chosen as a suitable systems of wavelets, then both Assumptions 3.2 and 3.3 are satisfied (see the discussion in Section 3.2 of [60]). This means that, in such cases, Theorem 3.4, as well as the other results from Section 3.2 of [60], are applicable, and yield the separation rate (as well as non-asymptotic results) for the problem of detecting signals that are linear combinations of certain wavelets.

Note however, that even if A is one of those operators, and $(u_k)_{k\in I}$ is chosen as an appropriate system of wavelets, the conditions of Theorem 3.1 are still not necessarily satisfied (see Lemma 3.7 of [60]).

3.4 Discussion

In this paper we extended the known results for minimax signal detection from indirect and noisy data. The types of alternatives that we considered here, have, to the best of our knowledge, only been studied for the case that the collection $(u_k)_{k\in I}$ (together with the collection $(v_k)_{k\in I}$ and the sequence $(\lambda_k)_{k\in I}$) forms the singular value decomposition (SVD) of the operator A. We think that this is interesting, since systems such as wavelets provide greater flexibility than the SVD.

However, these results should be taken with a grain of salt. First, we were not able to present a widely applicable criterion to decide when the conditions of Theorem 3.1 are satisfied. Second, given a collection $(u_k)_{k\in I}$, one could imagine other useful types of alternatives based on this collection, besides those two that we considered here (see for example the types of alternatives considered in [49]).

We refer to Section 5 of [60] for further discussion.

Connection to the setting from Chapter 2

Let us fix $n \in \mathbb{N}$ and $\sigma > 0$ and assume that we observe the random vector $Y_n = \mu_n + \sigma\xi_n$, with $\xi_n \sim \mathcal{N}(0, \Sigma_n)$, similar to the setting from Chapter 2. If the matrix Σ_n is symmetric and positive definite, then (by pre-whitening) we obtain

$$X_n := \Sigma_n^{-1/2} Y_n = \Sigma_n^{-1/2}\mu_n + \sigma\zeta_n,$$

where $\zeta_n \sim \mathcal{N}(0, \mathrm{id}_n)$. This is, in fact, an instance of the model (3.1). However, recall that asymptotic results in this chapter are given in terms of $\sigma \to 0$ (as opposed to $n \to \infty$). Nevertheless, the non-asymptotic results (see Theorems 3.9 and 3.10 of [60]) are applicable (if a suitable collection $(u_k)_{k\in I}$ that satisfies Assumption 3.2 is available).

Own contribution

I came up with the idea of this study and carried out the analysis by myself under the guidance of Frank Werner and Axel Munk. Furthermore, I performed the simulations, and wrote the manuscript by myself, supported through discussions and helpful comments by Frank Werner and Axel Munk.

Appendix

Appendix A.1

Bump detection in the presence of dependency: Does it ease or does it load?

Farida Enikeeva[1,2], Axel Munk[3,4,5], Markus Pohlmann[3] and Frank Werner[6]

[1]Laboratoire de Mathématiques et Applications, CNRS UMR 7348, Université de Poitiers, Site du Futuroscope - Téléport 2, 11 Boulevard Marie et Pierre Curie, Bâtiment H3 - TSA 61125, 86073 Poitiers Cedex 9, France
[2]The A. A. Kharkevich Institute for Information Transmission Problems, Russian Academy of Science, Bolshoy Karentny per. 19-1, 127051 Moscow, Russia
[3]Institute for Mathematical Stochastics, University of Göttingen, Goldschmidtstr. 7, 37077 Göttingen, Germany
[4]Felix-Bernstein-Institute for Mathematical Statistics in the Biosciences, University of Göttingen, Goldschmidtstr. 7, 37077 Göttingen, Germany
[5]Statistical Inverse Problems in Biophysics Group, Max Planck Institute for Biophysical Chemistry, Am Faßberg 11, 37077 Göttingen, Germany
[6]Institute for Mathematics, University of Würzburg, Emil-Fischer-Str. 30, 97074 Würzburg, Germany

Abstract

We provide the asymptotic minimax detection boundary for a bump, i.e. an abrupt change, in the mean function of a stationary Gaussian process. This will be characterized in terms of the asymptotic behavior of the bump length and height as well as the dependency structure of the process. A major finding is that the asymptotic minimax detection boundary is generically determined by the value of its spectral density at zero. Finally, our asymptotic analysis is complemented by non-asymptotic results for AR(p) processes and confirmed to serve as a good proxy for finite sample scenarios in a simulation study. Our proofs are based on laws of large numbers for non-independent and non-identically distributed arrays of random variables and the asymptotically sharp analysis of the precision matrix of the process.

Keywords: Minimax testing, time series, ARMA processes, weak laws of large numbers, change point detection, Toeplitz matrices.

1 Introduction

1.1 Model and problem statement

In this paper we consider observations of a triangular array of Gaussian vectors, $Y = \mu_n + \xi_n$, $n \in \mathbb{N}$ with the coordinates

$$Y_{i,n} = \mu_{i,n} + \xi_{i,n}, \qquad \xi_n = (\xi_{1,n}, \dots, \xi_{n,n})^T \sim \mathcal{N}_n(0, \Sigma_n), \tag{1.1}$$

with a known positive definite covariance matrix $\Sigma_n \in \mathbb{R}^{n\times n}$, but an unknown mean vector $\mu_n = (\mu_{1,n}, \dots, \mu_{n,n})^T \in \mathbb{R}^n$. We will furthermore assume that the noise ξ_n in (1.1) consists of n consecutive samples of a stationary process $(Z_t)_{t\in\mathbb{Z}}$.

For a proper asymptotic treatment, we will assume that μ_n is obtained from equidistantly sampling a function $m_n : [0,1] \to \mathbb{R}$ at sampling points $\frac{i}{n}, i = 1, \dots, n$, i.e. $\mu_n = \left(m_n\left(\frac{1}{n}\right), \dots, m_n\left(\frac{n}{n}\right)\right)^T$. Our goal is to analyze how difficult it is to detect abrupt changes of the function m_n based on the observations $Y = (Y_{1,n}, ..., Y_{n,n})^T$ coming from (1.1). Therefore, we focus on functions m_n of the form

$$m_n(x) = \begin{cases} \delta_n & \text{if } x \in I_n, \\ 0 & \text{else,} \end{cases} \tag{1.2}$$

i.e. m_n has a bump located at the interval $I_n \subset [0,1]$ of height $\delta_n \in \mathbb{R}$, see also Figure 1 for an illustration. We assume throughout the paper that the matrix Σ_n in (1.1) as well as the length of the bump $\lambda_n \in (0,1)$ are known, but that its amplitude δ_n and the exact position of the bump itself are unknown.

To formalize the detection problem, let us introduce some notation. For an interval $I \subset [0,1]$ we use $\mathbf{1}_I \in \mathbb{R}^n$ as abbreviation for the vector with entries

$$\mathbf{1}_I(i) = \begin{cases} 1 & \text{if } \frac{i}{n} \in I, \\ 0 & \text{else,} \end{cases} \qquad 1 \le i \le n.$$

Consequently, $\mu_n = \delta_n \mathbf{1}_{I_n}$ whenever m_n is of the form (1.2). Furthermore let

$$\mathcal{I} := \left\{ [a,b) \;\middle|\; 0 \le a < b \le 1 \right\}$$

be the set of all right-open intervals in $[0,1]$, and for a given length $\lambda \in (0,1)$ we introduce by

$$\mathcal{I}(\lambda) := \left\{ [a,b) \;\middle|\; 0 \le a < b \le 1, b-a = \lambda \right\}$$

the set of all right-open intervals in $[0,1]$ of length λ.

Now the problem to detect a bump of length λ_n in the signal μ_n from (1.1) can be understood as the hypothesis testing problem

$$\begin{gathered} H_0^n : Y \sim \mathcal{N}_n(0, \Sigma_n) \\ \text{against} \\ H_1^n : \exists I \in \mathcal{I}(\lambda_n),\ \exists \delta \in \mathbb{R} : |\delta| \ge \Delta_n \quad \text{such that} \quad Y \sim \mathcal{N}_n(\delta \mathbf{1}_I, \Sigma_n) \end{gathered} \tag{1.3}$$

with a minimal amplitude value $\Delta_n > 0$ to ensure distinguishability of H_0^n and H_1^n. Note that I and δ in (1.3) are allowed to depend on n (as the length λ_n and the minimal amplitude value Δ_n do), but we suppress this dependency in the following. Similarly we write H_0 instead of H_0^n as Σ_n is assumed to be known. Note that we will consider the situation $\lambda_n \to 0$ as $n \to \infty$ below, corresponding to a vanishing bump, which avoids trivial cases such as $\mathbb{E}_{H_1^n}[Y_i] = \delta > 0$ for all $1 \le i \le n$ in (1.3).

The aim of this paper is to provide insight on how the dependency structure in (1.1) encoded in terms of Σ_n influences the detection of such a bump. More precisely, we would like to derive

asymptotic conditions[1] on the minimal detectable bump amplitude Δ_n depending on Σ_n, λ_n and n. To the best of our knowledge, there is no systematic understanding of this problem from the minimax point of view. We will therefore provide (asymptotic) lower and upper bounds for the amplitude of asymptotically detectable signals in the following sense (cf. [38,40]). Let $\alpha, \beta \in (0,1)$ be arbitrary error levels.

upper detection bound: Whenever the bump amplitude Δ_n satisfies $\Delta_n = c\varphi_n$, $c \geq c^*$ with a constant $c^* > 0$ and a rate φ_n depending on n, λ_n and Σ_n, then **there is a sequence of tests** for (1.3) with (asymptotic) type I error $\leq \alpha$ and (asymptotic) type II error $\leq \beta$.

lower detection bound: Whenever the bump amplitude Δ_n satisfies $\Delta_n = c\tilde{\varphi}_n$, $c \leq c_*$ with a constant $c_* > 0$ and a rate $\tilde{\varphi}_n$ depending on n, λ_n and Σ_n, then **no sequence of tests** for (1.3) can have type (asymptotic) I error $\leq \alpha$ and at the same time (asymptotic) type II error $\leq \beta$.

Precise definitions of the (asymptotic) type I and type II errors and comments on the validity of these particular notions of the detection bounds will be given in Section 2.1. Note that the minimax separation rate φ_n might depend on the prescribed significance levels α and β, and that the definitions become trivial if $\beta \geq 1-\alpha$, as then any standard Bernoulli experiment with success probability α defines a corresponding test. However, in our case neither the constants c_* and c^* nor the rate depend on the error levels α and β. That is why in the following we will always choose $\alpha = \beta \in \left(0, \frac{1}{2}\right)$ and argue in Section 2.1 that this is sufficient.

If $\tilde{\varphi}_n = \varphi_n$ in the above upper and lower bound, then we speak of the **(asymptotic) minimax separation rate** $\Delta_n \sim \varphi_n$. If furthermore $c^* = c_*$, then $\Delta_n \asymp c_*\varphi_n = c^*\tilde{\varphi}_n$ is called the **(asymptotic) minimax detection boundary** over all possible amplitudes $\Delta_n > 0$ and positions $I \in \mathcal{I}(\lambda_n)$. We will provide explicit expressions for this under weak assumptions on the covariance matrix Σ_n.

We will provide lower and upper bounds in terms of sums over diagonal blocks within Σ_n (cf. Section 2.3 and Lemmas 8 and 9), and for the case of noise generated by subsequent samples of a stationary time series we will show that these lower and upper bounds coincide.

In case of i.i.d. observations, this is $\Sigma_n = \sigma^2 \mathrm{id}_n$ in (1.1), the minimax detection boundary is well-known and given by (see [14,19,27])

$$\Delta_n \asymp \sigma \sqrt{\frac{-2\log \lambda_n}{n\lambda_n}}. \tag{1.4}$$

Here, and in the following, we require

$$\lambda_n \to 0 \quad \text{and} \quad n\lambda_n \to \infty \qquad \text{as} \qquad n \to \infty. \tag{1.5}$$

Signals for which the left-hand side in (1.4) is asymptotically larger than the right-hand side can be detected consistently (in the sense of an upper detection bound as described above), whereas they can not be detected consistently once the left-hand side in (1.4) is asymptotically smaller than the right-hand side (in the sense of a lower detection bound as described above). Although (1.4) is known for a long time when the errors are i.i.d., to the best of our knowledge, the influence of the error dependency structure on the detection boundary (1.4) is an issue that is much less investigated systematically, although many methods to estimate such abrupt changes in the signal corrupted by serially dependent errors have been suggested (see Section 1.3). In this sense, this paper contributes a benchmark to such methods. Let us illustrate the effect of the dependency on (1.4) with ξ_n in (1.1) arising from an AR(1) process with unit variance and auto-correlation coefficient ρ, this is $\xi_n = \left(1-\rho^2\right)^{1/2}(Z_1, ..., Z_n)^T$ where $Z_t - \rho Z_{t-1} = \zeta_t$ with i.i.d. standard Gaussian noise $\zeta_t, t \in \mathbb{Z}$. In Figure 1 we illustrate three different situations

[1] Let $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ two sequences of positive numbers. In the following we write $a_n \sim b_n$ if $0 < \liminf_{n\to\infty} a_n/b_n \leq \limsup_{n\to\infty} a_n/b_n < \infty$, and $a_n \asymp b_n$ if $\lim_{n\to\infty} a_n/b_n = 1$.

encoded in terms of ρ, namely positively correlated noise ($\rho = 0.7$), independent noise ($\rho = 0$), and negatively correlated noise ($\rho = -0.7$). It seems intuitively clear that the value of ρ influences the difficulty of detecting a bump substantially, and especially positively correlated noise hinders efficient detection dramatically. Compare e.g. the first plot in Fig. 1, where noise and bump appear hardly to distinguish. Furthermore, due to the positive correlation, there appear several regions which suggest a bump in signal, which is not there. In contrast, the middle and bottom plot allow for simpler identification of the bump region. Our main result makes these intuitive findings precise.

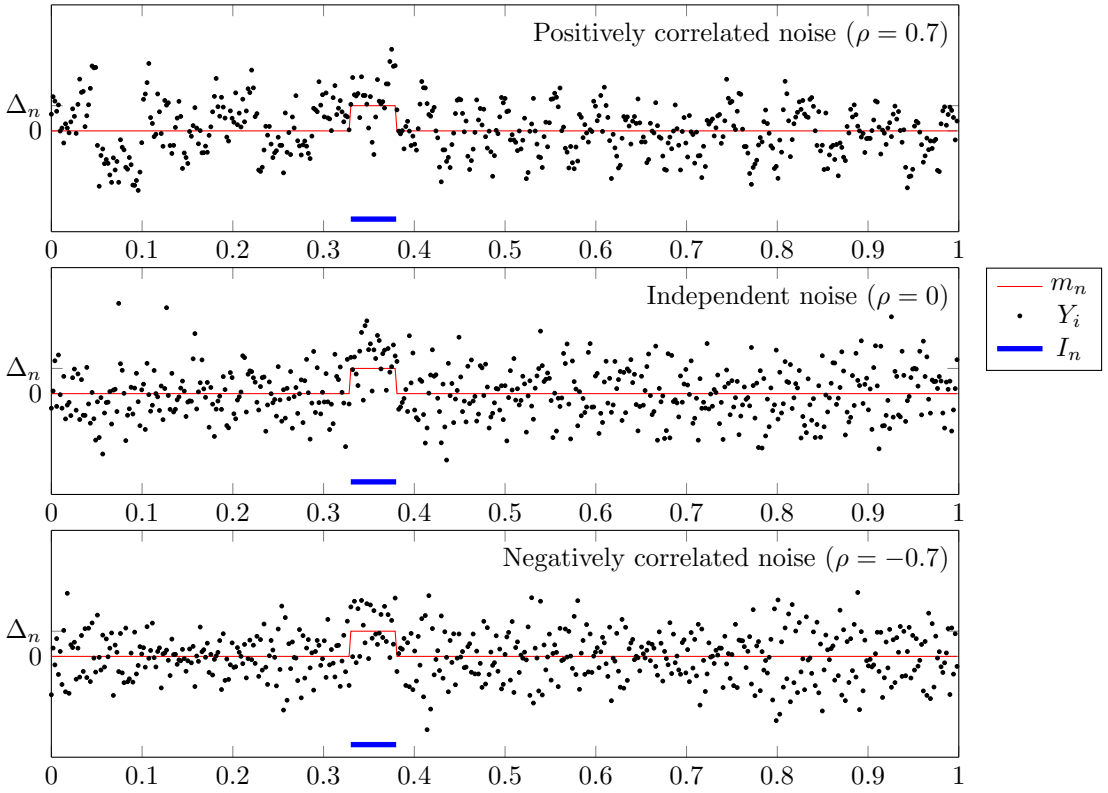


Figure 1: Model (1.1) in case of AR(1) noise for different values of ρ: Data together with the function m_n, where the model parameters are set to be $n = 512$ and $\Delta_n = 1$, $\sigma = 1$.

1.2 Results

To describe our results concerning the detection boundary for serially dependent data we require some more terminology. Let the autocovariance function γ_Z of the stationary process $(Z_t)_{t\in\mathbb{Z}}$ be given by $\gamma_Z(h) = \mathrm{Cov}\,[Z_t, Z_{t+h}]$ for $h \in \mathbb{Z}$. Assume that γ_Z is square summable, then the process Z has the spectral density $f_Z \in \mathbf{L}_2\,[-1/2, 1/2)$ defined by

$$f_Z(\nu) = \sum_{h=-\infty}^{\infty} \gamma_Z(h) e^{-2\pi i h \nu}, \qquad \nu \in [-1/2, 1/2)\,.$$

In fact, f_Z can also be considered as a function on the unit sphere, i.e. one naturally has $\lim_{\nu\to 1/2} f_Z(\nu) = f_Z(-1/2)$. We will also assume that the autocovariance function is symmetric, which is equivalent to f_Z being real-valued. In the following, we will omit the subscript Z in the notation of the spectral density of Z when it does not create ambiguities.

With this notation introduced, we will show under mild conditions that the detection boundary for the hypothesis testing problem (1.3) is given by

$$\Delta_n \asymp \sqrt{\frac{-2f\left(0\right)\log\lambda_n}{n\lambda_n}}.$$

It is immediately clear, that in case of independent observations where $\Sigma_n = \sigma^2 \mathrm{id}_n$, one has $f\left(0\right) = \sigma^2$, which reproduces (1.4). In the general case, note that

$$f\left(0\right) = \sum_{h\in\mathbb{Z}} \gamma\left(h\right),$$

i.e. the detection boundary solely depends on the value of the spectral density at zero which is known as *long-run variance*.[2]

In case of the AR(1)-based noise $\xi_n := (1-\rho^2)^{1/2}(Z_1,\ldots,Z_n)^T$ with unit variance as shown in Figure 1, the auto-covariance of the underlying AR(1) process Z_t is given by $\gamma_Z\left(h\right) = \gamma_Z(0)\rho^{|h|}$, where $\gamma_Z(0) = (1-\rho^2)^{-1}$. Thus the spectral density at zero of the noise process $\xi = \left((1-\rho^2)^{1/2}Z_i\right)_{i\in\mathbb{N}}$ is

$$f_\xi\left(0\right) = \left(1-\rho^2\right)\sum_{h=-\infty}^{\infty}\gamma_Z(h) = \frac{1+\rho}{1-\rho},$$

and hence the detection boundary is given by

$$\Delta_n \asymp \sqrt{\frac{1+\rho}{1-\rho}}\sqrt{\frac{-2\log\lambda_n}{n\lambda_n}}. \tag{1.6}$$

As an immediate consequence, this shows that bump detection is easier under a negative correlation ρ than in case of positive correlations. For the three values employed in Figure 1 we compute for the factor $\sqrt{\frac{1+\rho}{1-\rho}}$ in (1.6) the values 2.38 when $\rho = 0.7$ and 0.42 when $\rho = -0.7$. This means that the amplitude of detectable signals for $\rho = 0.7$ and $\rho = -0.7$ differs approximately by a factor of 5.6. Also, given the bump length λ_n, the detection of a bump of the same size Δ_n for $\rho = 0.7$ requires approximately a 6 times larger sample size than for $\rho = 0$, and even a 31 times larger sample size than for $\rho = -0.7$. This is in good agreement with the intuitive findings from Figure 1 and confirmed in finite sample situations in Section 4. In the simulations we also investigate the influence of several bumps instead of one, and find that independent of ρ, multiple bumps always help detection, as to be expected.

Remarkably, as in the case of i.i.d. noise with variance σ^2, where we have $f(0) = \sigma^2$, certain dependent error processes might also satisfy $f(0) = \sigma^2$, and hence obey the same difficulty to detect a bump as for the independent case. As an example, consider the stationary and causal AR(2) process given by $Z_t = \frac{1}{2}Z_{t-1} - \frac{1}{2}Z_{t-2} + \zeta_t$, where $\zeta_t \sim \mathcal{N}(0,1)$ for $t \in \mathbb{Z}$. In this case $f_Z\left(0\right) = \frac{1}{2} - \frac{1}{2} + 1 = 1$, even though the process Z_t is clearly not independent (see Section 3 for a comprehensive treatment of ARMA processes).

Proof strategy. To prove a lower detection bound, we will employ techniques dating back to Ingster [38] and Dümbgen and Spokoiny [19] developed for independent observations. To generalize this approach to our dependent case, we will use a recent weak law of large numbers due to Wang and Hu [71] for triangular arrays of random variables that are non-independent within each row and non-identically distributed between rows (see Section 6.1 for the precise statement and also [30, 57, 62, 67] for related results).

For the upper detection bound, we will provide an explicit test based on the supremum of the moving average process $\left(\mathbf{1}_I^T Y\right)_{I\in\mathcal{I}(\lambda_n)}$. A valid critical value will be given based on a chaining

[2]The long-run variance of a process $(Z_t)_{t\in\mathbb{Z}}$ with spectral density f is defined as $\lim_{n\to\infty} n^{-1}\mathrm{Var}\left[S_n\right]$, where $S_n = \sum_{i=1}^{n} Z_i$. It holds that $\lim_{n\to\infty} n^{-1}\mathrm{Var}\left[S_n\right] = f(0)$ (see [36] and Section 6.1 for details).

technique. Note that this cannot be obtained by a continuous upper bound of the stochastic process (as e.g. provided in Theorem 6.1 in [19]) due to the fact that the dependency structure is allowed to change with n and hence there is no continuous analog of $\left(\mathbf{1}_I^T Y\right)_{I\in\mathcal{I}(\lambda_n)}$.

1.3 Related work

Bump detection for dependent data appears to be relevant to a variety of applications where piecewise constant signals (i.e. several bumps) are observed under dependent noise. Exemplary, we mention molecular dynamics (MD) simulations, where collective motion characteristics of protein atoms are studied over time (see e.g. [47] and the references therein). For certain proteins it has been shown that the noise process can be well modeled by a stationary ARMA(p,q) process with small p and q, see [66]. Another application is the analysis of ion channel recordings, where one aims to identify opening and closing states of physiologically relevant channels (see [56] and the references therein). Here, the dependency structure is induced by a known band-pass filter, ensuring that Σ_n in (1.1) can be precomputed explicitly (which corresponds to our setting of known Σ_n), and allowing for a good approximation by stationary and m-dependent noise with small m, see [59].

In fact, bump detection as discussed here is closely related to estimation of a signal which consists of piece-wise constant segments, often denoted as change point *estimation*. We refer to the classical works of Ibragimov and Has'minskii [35], Csörgő and Horváth [15], Brodsky and Darkhovsky [11], Carlstein, Müller and Siegmund [12], and Siegmund [65] for a survey of the existing results as well as to the review article by Aue and Horváth [3]. Indeed, if the bumps have been properly identified by a detection method, posterior estimation of the signal is relatively easy, see [27] for such a combined approach in case of i.i.d. errors, and [17] in case of dependent data. We also mention [13], who presented a robust approach for AR(1) errors.

Model (1.1) can be seen as prototypical for the more complex situation when several bumps are to be detected. We do not intend to provide novel methodology for this situation in this paper, rather Theorem 1 provides a benchmark for *detecting* such a bump which then can be used to benchmark the detection power of any method designed for this task. Minimax detection has a long history, see e.g. the seminal series of papers by Ingster [38] or the monograph by Tsybakov [69]. More recently, Goldenshluger, Juditsky and Nemirovski [29] provided a general approach based on convex optimization. In case of independent observations, the problem of detecting a bump has been considered in [5,9,14,20,27,42], and our strategy of proof for the lower bound is adopted from [19]. We also mention [25] for a model with a simultaneous bump in the variance, and [58] for heterogeneous noise, however still restricted to independent observations.

The literature on minimax detection for dependent noise is much less developed, and most similar in spirit to our work are the papers by Hall and Jin [31] and Keshavarz, Scott and Nguyen [44]. In the former, the minimax detection boundary for an unstructured version of the model (1.1) in a Bayesian setting is derived, that is $\mathbb{P}\left[m_n\left(\frac{i}{n}\right)=\Delta_n\right]=\rho_n$ and $\mathbb{P}\left[m_n\left(\frac{i}{n}\right)=0\right]=1-\rho_n$ with a probability ρ_n tending to 0. In contrast to [31], in the present setting we can borrow strength from neighboring observations in a bump. Still, we can exploit a result in [31] about the decay behavior of inverses of covariance matrices (see Section 6.1) to validate Assumption 2. Keshavarz, Scott and Nguyen [44] deal with the classical change-point in mean problem, i.e. with the problem to detect whether $m_n(i/n)\equiv 0$ for all $1\le i\le n$, or if there exists $\tau\in[1,n]$ such that $m_n(i/n)=-\frac{1}{2}\Delta_n\mathbf{1}\{i\le\tau\}+\frac{1}{2}\Delta_n\mathbf{1}\{i>\tau\}$ for $1\le i\le n$. The authors derive upper and lower bounds for detection from dependent data as in (1.1), similar in spirit to our Theorem 1. Their bounds, however, do not coincide with ours, i.e. they do not derive the precise minimax detection boundary, as they are mostly interested in the rate of estimation. However, as we see from Theorem 1, the $\sqrt{-\log\lambda_n}$ rate does not change, it is the constant $f(0)$ which matters. We will employ several of their computations concerning covariance structures of time series (while correcting a couple of technical inaccuracies).

We finally comment on the assumption of knowing Σ_n and the length λ_n. If λ_n is unknown, estimation of the function m_n can be performed in the independent noise case by SMUCE [27] via

a multiscale approach. SMUCE is known to achieve the asymptotic detection boundary (1.4) in case of i.i.d. Gaussian errors. For the dependent case with a (partially) unknown covariance matrix Σ_n, further methods for estimation of m_n such as H-SMUCE [58], J-SMURF [33] or JULES [59] have been developed. They all rely on a local estimation of the covariance structure in combination with a multiscale approach. None of these methods achieves the detection boundary derived in this paper, and hence it remains unclear if not knowing Σ_n and / or λ_n would affect it. Developing a test which achieves a corresponding upper bound by multiscale methods is beyond the scope of this paper and is postponed to future work.

1.4 Organization of the paper

The remaining part of this paper is organized as follows: In Section 2 we give a precise statement of our assumptions and formulate our main theorem. Also non-asymptotic results are discussed here. The implications for ARMA models are then given in Section 3, where the previously mentioned non-asymptotic results are specified for AR(p) noise. In Section 4 we present some simulations which support that our asymptotic theory is already useful for small samples. All proofs are deferred to Section 5 and Section 6 in Supplement A.2.

2 Main results

2.1 Notation and assumptions

To treat the testing problem (1.3), we will consider tests $\Phi_n : \mathbb{R}^n \to \{0,1\}$, $n \in \mathbb{N}$, where $\Phi_n(Y) = 0$ means that the null hypothesis H_0 is accepted, and $\Phi_n(Y) = 1$ means that the null hypothesis is rejected, i.e. the presence of a bump is concluded.

Denote by $\mathbb{P}_0$ the measure $\mathcal{N}_n(0, \Sigma_n)$ of Y under the null hypothesis and by $\mathbb{P}_{I,\delta}$ the measure $\mathcal{N}_n(\delta \mathbf{1}_I, \Sigma_n)$ of Y given that there is a bump of height δ within the interval I. With this we will denote the corresponding expectations accordingly by $\mathbb{E}_0$ and $\mathbb{E}_{I,\delta}$. We define the type I error of Φ_n by

$$\bar{\alpha}(\Phi_n, \Sigma_n) := \mathbb{E}_0[\Phi_n(Y)] = \mathbb{P}_0[\Phi_n(Y) = 1].$$

Furthermore, we say that a sequence $(\Phi_n)_{n\in\mathbb{N}}$ of such tests has asymptotic level $\alpha \in [0,1]$ if $\limsup_{n\to\infty} \bar{\alpha}(\Phi_n, \Sigma_n) \leq \alpha$. The type II error depending on the parameters Σ_n, Δ_n and λ_n is defined as

$$\bar{\beta}(\Phi_n, \Sigma_n, \Delta_n, \lambda_n) := \sup_{I\in\mathcal{I}(\lambda_n)} \sup_{|\delta|\geq\Delta_n} \mathbb{P}_{I,\delta}[\Phi_n(Y) = 0].$$

For a sequence $(\Phi_n)_{n\in\mathbb{N}}$ of such tests we define its asymptotic type II error to be $\limsup_{n\to\infty} \bar{\beta}(\Phi_n, \Sigma_n, \Delta_n, \lambda_n)$. The asymptotic power of such a family is then given by $1 - \limsup_{n\to\infty} \bar{\beta}(\Phi_n, \Sigma_n, \Delta_n, \lambda_n)$. For the sake of brevity, we might suppress the dependency on the parameters in the following and write only $\bar{\alpha}(\Phi_n)$ and $\bar{\beta}(\Phi_n)$, respectively.

With this notation, we can now precisely recall the requirements for lower and upper bounds on detectability as discussed in the introduction:

upper detection bound: For any $\alpha \in \left(0, \frac{1}{2}\right)$, there exist $c^* > 0$ and a sequence of tests $\Phi^*_{n,\alpha}$, $n \in \mathbb{N}$ of asymptotic level α such that $\forall c > c^*$,

$$\limsup_{n\to\infty} \bar{\beta}(\Phi_n, \Sigma_n, c\varphi_n, \lambda_n) \leq \alpha.$$

Note that this notion of the upper detection bound is in accordance with the usual minimax testing paradigm (cf. Ingster and Suslina [40]), as it implies that

$$\lim_{n\to\infty} \inf_{\Phi\in\Psi_n} [\bar{\alpha}(\Phi, \Sigma_n) + \bar{\beta}(\Phi, \Sigma_n, c\varphi_n, \lambda_n)] = 0,$$

as $n \to \infty$, since α was arbitrary. Here, Ψ_n is the collection of all tests for the testing problem (1.3) given n observations.

lower detection bound: For any $\alpha \in \left(0, \frac{1}{2}\right)$, there exists $c_* > 0$ such that $\forall c < c_*$, and for any sequence of tests Φ_n, $n \in \mathbb{N}$ of asymptotic level α,

$$\liminf_{n\to\infty} \bar{\beta}\left(\Phi_n, \Sigma_n, c\tilde{\varphi}_n, \lambda_n\right) \geq 1 - \alpha.$$

This implies that

$$\lim_{n\to\infty} \inf_{\Phi\in\Psi_n} \left[\bar{\alpha}\left(\Phi, \Sigma_n\right) + \bar{\beta}\left(\Phi, \Sigma_n, c\varphi_n, \lambda_n\right)\right] = 1.$$

The choice of $1 - \alpha$ as the lower bound of the limit of the type II errors in the lower detection bound is justified by the fact that the minimax testing risk is bounded from below as follows (see [40], p. 55, Theorem 2.1):

$$\inf_{\Phi\in\Psi} \left[\bar{\alpha}\left(\Phi, \Sigma_n\right) + \bar{\beta}\left(\Phi, \Sigma_n, c\varphi_n, \lambda_n\right)\right] \geq 1 - \frac{1}{2}\|[\mathcal{P}_0], [\mathcal{P}_1]\|_1,$$

where $\|[\mathcal{P}_0], [\mathcal{P}_1]\|_1$ is the L_1-distance between the convex hulls of measures corresponding to the null and the alternative hypotheses and Ψ is the set of all possible tests. It implies that the type II error of the α-level test will be always greater or equal $1 - \alpha$ for non-distinguishable null and alternative hypotheses.

To derive lower and upper bounds in this sense, we will now pose some assumptions on the possible lengths λ_n of intervals and the covariance structure Σ_n:

Assumption 1. *We assume that*

(i) $\frac{n\lambda_n}{\log n} \to \infty$ *as* $n \to \infty$,

(ii) $\lambda_n = o\left(\frac{1}{\log n}\right)$ *as* $n \to \infty$.

The first part of Assumption 1 assures that the number of observations within any interval of length λ_n is at least of logarithmic order as $n \to \infty$. The second condition of Assumption 1, however, gives a bound for the maximal length of the considered intervals, which ensures less than $n/\log n$ observations in the bump interval. Roughly speaking both conditions are required to have enough complementary observations (outside respectively inside the bump) to guarantee asymptotic detection. Note that, in particular, Assumption 1(ii) means that $\lambda_n \to 0$ as $n \to \infty$, i.e. Assumption 1 especially implies (1.5). We emphasize that conditions as in (ii) restricting λ_n from being too large are common. Assumption 1 plays a crucial role in the proof of the upper bound, whereas the lower bound can be established under milder conditions (1.5).

However, note that when we consider a slightly modified version of the testing problem (1.3) where the bump may not occur in any interval of length λ_n, but only within a candidate set $I_k := [(k-1)\lambda_n, k\lambda_n)$, $1 \leq k \leq \lfloor 1/\lambda_n \rfloor$ of non-overlapping intervals, then Assumption 1 can be replaced by (1.5) and the detection boundary will remain the same (cf. Section 2.3).

Instead of posing assumptions on Σ_n directly, we will again employ the spectral density f of the underlying stationary process Z as mentioned in the introduction. To do so, we require some more terminology. For a function $g \in \mathbf{L}_2\left[-1/2, 1/2\right)$, we denote by $\mathcal{T}(g)$ the Toeplitz matrix generated by g, i.e. the matrix with entries $\left(\mathcal{T}(g)\right)_{i,j\in\mathbb{N}} = g_{j-i}$, where

$$g_k = \int_{-\frac{1}{2}}^{\frac{1}{2}} g(u)e^{-2\pi iku}du, \qquad k \in \mathbb{Z},$$

is the k-th Fourier coefficient of g. Note that this allows us to encode the covariance matrix Σ_n completely in terms of f. More precisely, the covariance matrix Σ_n of the noise ξ_n in (1.1) has entries $\Sigma_n(i,j) = \gamma(|i-j|) = f_{|i-j|}$, and we see that $\Sigma_n =: \mathcal{T}_n(f)$ is the n-th truncated Toeplitz matrix generated by f, i.e. the upper left $n \times n$ submatrix, of $\mathcal{T}(f)$. Consequently, we will also pose the corresponding assumptions in terms of the function f, which allows us to derive results for any sequence $\left(\Sigma_n\right)_{n\geq 1}$ of covariance matrices which are generated by such an f (and not only for specific dependent processes):

Assumption 2. *Let $(\Sigma_n)_{n\geq 1}$ be a sequence of covariance matrices such that $\Sigma_n = \mathcal{T}_n(f)$ as introduced above with a function $f : [-1/2, 1/2) \to \mathbb{R}$, that is continuous and satisfies $\lim_{\nu\to 1/2} f(\nu) = f(-1/2)$ and $\operatorname{essinf}_{\nu\in[-1/2,1/2)} f(\nu) > 0$. Further, suppose that the Fourier coefficients f_h, $h \in \mathbb{Z}$ of f decay sufficiently fast, i.e. there are constants $C > 0$ and $\kappa > 0$, such that*

$$|f_h| \leq C(1 + |h|)^{-(1+\kappa)}, \quad h \in \mathbb{Z}.$$

Assumption 2 ensures that the dependency between $\mathbf{1}_I^T Y$ and $\mathbf{1}_{I'}^T Y$ for two candidate intervals $I, I' \in \mathcal{I}(\lambda_n)$, will be asymptotically small as soon as they are disjoint. It excludes trivial cases such as total dependence described by $\Sigma_n(i, j) = 1$ for all $i, j \in \{1, ..., n\}$, but also permits spectral densities f with only slowly decaying Fourier coefficients such as discontinuous functions.

Note that also sequences of covariance matrices of the form $(\Sigma_n)_{i,j} = g\left(\frac{|i-j|}{n}\right)$, $1 \leq i, j \leq n$, $n \in \mathbb{N}$, where g is some kernel function, are prohibited due to this assumption. Covariance matrices of this kind would have the undesired effect to make the dependency between $\mathbf{1}_I^T Y$ and $\mathbf{1}_{I'}^T Y$ even for disjoint candidate intervals $I, I' \in \mathcal{I}(\lambda_n)$ stronger as the length λ_n vanishes.

2.2 Asymptotic detection boundary

Our main theorem will be the following.

Theorem 1. *If Assumptions 1 and 2 hold for the bump regression model* (1.1)*, then the asymptotic minimax detection boundary for the testing problem* (1.3) *is given by*

$$\Delta_n \asymp \sqrt{\frac{-2f(0)\log\lambda_n}{n\lambda_n}},$$

as $n \to \infty$.

For the details of the proof we refer to Section 5. The upper bound will be achieved by a specific test Φ_n^{a}, which scans over all intervals of length λ_n, given by

$$\Phi_n^{\mathrm{a}}(Y) = \begin{cases} 1 & \text{if } \sup_{I\in\mathcal{I}(\lambda_n)} \frac{|\mathbf{1}_I^T Y|}{\sqrt{\mathbf{1}_I^T \Sigma_n \mathbf{1}_I}} > c_{\alpha,n}, \\ 0 & \text{else,} \end{cases} \tag{2.1}$$

where the threshold $c_{\alpha,n}$ will be determined in the proof of Lemma 8 in Section 5. Note that this test is not a likelihood ratio type test (as the LRT relies on $\mathbf{1}_I^T \Sigma_n^{-1} Y$ instead of $\mathbf{1}_I^T Y$).

For the proof of the lower bound we employ a strategy from Dümbgen and Spokoiny [19], and use a very specific law of large numbers for arrays of non-independent and non-identically distributed random variables.

2.3 Non-asymptotic results

Note that Theorem 1 yields only an asymptotic result. In this section we give non-asymptotic results in the case of a seemingly simpler testing problem with possible bumps that belong to a set of non-overlapping intervals. This is formalized by considering the set $\mathcal{I}^0$ of non-overlapping candidate intervals given by

$$\mathcal{I}^0 := \left\{ I_k \;\middle|\; 1 \leq k \leq \lfloor \lambda_n^{-1} \rfloor \right\}, \qquad I_k := \big[(k-1)\lambda_n, k\lambda_n\big), \quad 1 \leq k \leq \lfloor \lambda_n^{-1} \rfloor. \tag{2.2}$$

The goal is still to detect the presence of the bump (but with position being only in $\mathcal{I}^0$) and to derive non-asymptotic results on the detection boundary for the testing problem

$$\begin{gathered} H_0 : Y \sim \mathcal{N}_n(0, \Sigma_n) \\ \text{against} \\ H_1^n : \exists 1 \leq k \leq \lfloor \lambda_n^{-1} \rfloor,\ \exists \delta \in \mathbb{R} : |\delta| \geq \Delta_n \quad \text{such that} \quad Y \sim \mathcal{N}_n(\delta \mathbf{1}_{I_k}, \Sigma_n) \end{gathered} \tag{2.3}$$

Note that this testing problem might seem much simpler than (1.3) at a first glance, but we will see, however, that the (asymptotic) detection boundary is in fact the same. Concerning the lower bound, this can be seen readily from the proof of Theorem 1, cf. Lemma 9.

To detect a bump, we will here employ the maximum likelihood ratio test

$$\Phi_n^{\mathrm{d}}(Y) = \mathbf{1}\left\{T_n^0(Y) > c_{\alpha,n}\right\}$$

based on the statistic

$$T_n^0(Y) = \sup_{I\in\mathcal{I}^0} \frac{|\mathbf{1}_I^T\Sigma_n^{-1}Y|}{\sqrt{\mathbf{1}_I^T\Sigma_n^{-1}\mathbf{1}_I}} = \sup_{1\le k\le \lfloor\lambda_n^{-1}\rfloor} \frac{|\mathbf{1}_{I_k}^T\Sigma_n^{-1}Y|}{\sqrt{\tilde\sigma_k}}, \tag{2.4}$$

where we denote

$$\tilde\sigma_k = \mathbf{1}_{I_k}^T\Sigma_n^{-1}\mathbf{1}_{I_k}, \quad k = 1,\dots,\lfloor\lambda_n^{-1}\rfloor. \tag{2.5}$$

The quantities $\tilde\sigma_k$ are in fact the variances of $\mathbf{1}_I^T\Sigma_n^{-1}Y$ corresponding to the sum of $\lfloor n\lambda_n\rfloor$ random variables with covariance structure given by the I_k-block of Σ_n^{-1}. The type I and II errors of the test Φ_n^{d} are defined as

$$\tilde\alpha\left(\Phi_n^{\mathrm{d}}\right) := \mathbb{P}_0\left[\Phi_n^{\mathrm{d}}(Y) = 1\right] \quad \text{and} \quad \tilde\beta\left(\Phi_n^{\mathrm{d}}\right) := \sup_{I\in\mathcal{I}^0}\sup_{|\delta|\ge\Delta_n} \mathbb{P}_{I,\delta}\left[\Phi_n^{\mathrm{d}}(Y) = 0\right].$$

Then the following result establishes basic properties of the test Φ_n^{d}.

Theorem 2. *Consider the testing problem (2.3) and let $\alpha\in(0,1)$ be any fixed significance level. For the maximum likelihood ratio test Φ_n^{d} set*

$$c_{\alpha,n} := \sqrt{2\log\frac{2}{\alpha\lambda_n}}. \tag{2.6}$$

Then it holds $\tilde\alpha\left(\Phi_n^{\mathrm{d}}\right)\le\alpha$ for all $n\in\mathbb{N}$ and

$$\tilde\beta\left(\Phi_n^{\mathrm{d}}\right) \le \mathbb{P}\left[|Z| > \Delta_n \inf_{1\le k\le\lfloor\lambda_n^{-1}\rfloor}\sqrt{\tilde\sigma_k} - c_{\alpha,n}\right]$$

with a standard Gaussian random variable $Z\sim\mathcal{N}(0,1)$ and $\tilde\sigma_k$ as in (2.5).

The proof is obtained by straightforward computations, see Section 6.2 of Supplement A.2 for details. Theorem 2 yields explicit non-asymptotic bounds for the test Φ_n^{d}, but those do also yield an asymptotic upper bound for the detection boundary:

Corollary 3. *Let $(\varepsilon_n)_{n\in\mathbb{N}}$ be a positive sequence satisfying*

$$\varepsilon_n\sqrt{-\log\lambda_n} \ge \sqrt{\log\frac{2}{\alpha}} + \sqrt{\log\frac{1}{\alpha}}, \tag{2.7}$$

and suppose that the bump altitude Δ_n in the testing problem (2.3) obeys

$$\Delta_n \inf_{1\le k\le\lfloor\lambda_n^{-1}\rfloor}\sqrt{\tilde\sigma_k} \ge \sqrt{2}\,(1+\varepsilon_n)\sqrt{-\log\lambda_n}. \tag{2.8}$$

Then the asymptotic type II error of Φ_n^{d} with $c_{\alpha,n}$ as in (2.6) satisfies

$$\limsup_{n\to\infty}\tilde\beta\left(\Phi_n^{\mathrm{d}}\right)\le\alpha,$$

This shows that the upper bound to be obtained by Φ_n^{d} depends only on the asymptotic behavior of $\inf_{1\le k\le\lfloor\lambda_n^{-1}\rfloor}\sqrt{\tilde\sigma_k}$ with $\tilde\sigma_k$ as in (2.5). Inspecting the proof of Lemma 9, we find that we can derive an according lower bound depending only on the asymptotic behavior of $\sup_{1\le k\le\lfloor\frac{1}{\lambda_n}\rfloor}\sqrt{\tilde\sigma_k}$. In case of AR($p$) noise we will see in Section 2.3 that these quantities can be computed explicitly and will asymptotically equal in agreement with Theorem 1.

3 ARMA processes and finite sample results

3.1 Application to ARMA processes

Suppose that the noise vector $\xi_n = (Z_1, \dots, Z_n)^T$ in (1.1) is sampled from n consecutive realizations of a stationary ARMA(p, q) time series Z_t, with $p \geq 0$, $q \geq 0$ defined as

$$\phi(B)Z_t = \theta(B)\zeta_t, \quad \zeta_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1), \quad t \in \mathbb{Z}. \tag{3.1}$$

Here B is the so-called backshift operator, defined by $BX_t = X_{t-1}$, and $\phi(z)$ and $\theta(z)$, $z \in \mathbb{C}$, are polynomials of degrees p and q, respectively, given by

$$\phi(z) = 1 + \sum_{i=1}^{p} \phi_i z^i, \quad \theta(z) = 1 + \sum_{i=1}^{q} \theta_i z^i. \tag{3.2}$$

We further suppose that ϕ and θ have no common roots, and that all roots of both ϕ and θ lie outside of the unit circle $\{z \in \mathbb{C} : |z| \leq 1\}$ (see [10] for more details).

Denote by γ the auto-covariance function of Z, i.e. $\gamma(h) = \mathbb{E}[Z_t Z_{t+h}]$ for $h \in \mathbb{Z}$ (as clearly $\mathbb{E}[Z_t] = 0$ for all $t \in \mathbb{Z}$). It is well-known (see for example [10], Theorem 4.4.2), that in the case of an ARMA(p, q) time series, its spectral density is given by

$$f(\nu) = \frac{|\theta(e^{-2\pi i\nu})|^2}{|\phi(e^{-2\pi i\nu})|^2}, \quad \nu \in [-1/2, 1/2). \tag{3.3}$$

Note that the spectral density f is continuous at 0 as well as the function $1/f$, since the process is reversible and causal under the posed assumptions on ϕ and θ. Thus, applying Theorem 1 to this setting immediately yields the following:

Theorem 4. *Assume that we are given observations from* (1.1)*, where the noise ξ_n is given by n consecutive samples of an ARMA(p, q) time series as in* (3.1) *with the polynomials ϕ and θ in* (3.2) *having no common roots and no roots within the unit circle. Furthermore, assume that Assumption 1 holds. Then the asymptotic detection boundary of the hypothesis testing problem* (1.3) *is given by*

$$\Delta_n \asymp \sqrt{\frac{-2f(0)\log\lambda_n}{n\lambda_n}} = \left|\frac{1 + \sum_{i=1}^{q}\theta_i}{1 + \sum_{i=1}^{p}\phi_i}\right| \sqrt{\frac{-2\log\lambda_n}{n\lambda_n}}, \tag{3.4}$$

as $n \to \infty$.

We find that the presence of dependency either eases or loads the bump detection, depending on $f(0) = |\theta(1)/\phi(1)|^2$ (which is 1 in the independent noise case). If $f(0) < 1$, then the detection becomes simpler (and smaller bumps are still consistently detectable), but if $f(0) > 1$ detection becomes more difficult. For AR(1) noise, this issue was already discussed in the introduction.

3.2 Non-asymptotic results for AR(p)

In this Section we will derive non-asymptotic results for the specific case of AR(p) noise. Let us therefore specify (3.1) to a stationary AR(p) process Z_t,

$$\sum_{i=0}^{p} \phi_i Z_{t-i} = \zeta_t, \quad t \in \mathbb{Z} \tag{3.5}$$

with independent standard Gaussian innovations ζ_t. In the notation of (3.1), we have $\phi(z) = \sum_{i=0}^{p} \phi_i z^i$ and $\theta(z) \equiv 1$. Again, we work under the standard assumptions that the characteristic polynomial $\phi(z)$ has no zeros inside the unit circle $\{z \in \mathbb{C} : |z| \leq 1\}$. Note that in this case $f(0) = \left|\sum_{i=0}^{p} \phi_i\right|^{-2}$.

We have seen in the discussion succeeding Theorem 2 that the upper and lower bounds depend on the quantities $\tilde{\sigma}_k = \mathbf{1}_{I_k}^T \Sigma_n^{-1} \mathbf{1}_{I_k}$ and correspondingly, their minimal and maximal values. Theorem 4 gives the detection boundary condition for ARMA noise with an asymptotic risk constant. Since $\tilde{\sigma}_k$ is just the sum over the block of Σ_n^{-1}, using the exact inverse of Σ_n (see the appendix for the exact formula of Σ_n^{-1} obtained by Siddiqui [64]), we can calculate the minimax risk constants exactly.

Lemma 5. *Let Σ_n be the auto-covariance matrix induced by an AR(p) process Z_t and $\tilde{\sigma}_k = \mathbf{1}_{I_k}^T \Sigma_n^{-1} \mathbf{1}_{I_k}$, $k = 1, \ldots, \lfloor \lambda_n^{-1} \rfloor$. Assume that $1 \leq \lfloor n\lambda_n \rfloor \leq n - 2p$ and $n > 3p$.*

1. *If $\lfloor n\lambda_n \rfloor \leq p$, then*

$$\inf_{1 \leq k \leq \lfloor \lambda_n^{-1} \rfloor} \tilde{\sigma}_k = \sum_{i=1}^{\lfloor n\lambda_n \rfloor} \left(\sum_{t=0}^{i-1} \phi_t \right)^2, \tag{3.6}$$

$$\sup_{1 \leq k \leq \lfloor \lambda_n^{-1} \rfloor} \tilde{\sigma}_k = \inf_{1 \leq k \leq \lfloor \lambda_n^{-1} \rfloor} \tilde{\sigma}_k + \sum_{i=0}^{p - \lfloor n\lambda_n \rfloor} \left(\sum_{t=1}^{\lfloor n\lambda_n \rfloor} \phi_{t+i} \right)^2 + \sum_{i=p-\lfloor n\lambda_n \rfloor}^{p} \left(\sum_{t=0}^{p-i} \phi_{t+i} \right)^2. \tag{3.7}$$

2. *If $p < \lfloor n\lambda_n \rfloor \leq n - 2p$, then*

$$\inf_{1 \leq k \leq \lfloor \lambda_n^{-1} \rfloor} \tilde{\sigma}_k = (\lfloor n\lambda_n \rfloor - p) \left(\sum_{t=0}^{p} \phi_t \right)^2 + \sum_{i=1}^{p} \left(\sum_{t=0}^{i-1} \phi_t \right)^2, \tag{3.8}$$

$$\sup_{1 \leq k \leq \lfloor \lambda_n^{-1} \rfloor} \tilde{\sigma}_k = \inf_{1 \leq k \leq \lfloor \lambda_n^{-1} \rfloor} \tilde{\sigma}_k + \sum_{i=1}^{p} \left(\sum_{t=0}^{p-i} \phi_{t+i} \right)^2. \tag{3.9}$$

We can now use the results of Theorem 2 and get the exact detection boundaries for two different regimes, when $\lfloor n\lambda_n \rfloor \leq p$ and $p < \lfloor n\lambda_n \rfloor \leq n - 2p$. Note that condition (5.4) is automatically satisfied since the inverse covariance matrix Σ_n^{-1} is $2p + 1$-diagonal.

Corollary 6. *Assume that possible locations k of the bump $I_k \in \mathcal{I}^0$ are separated from the endpoints of the interval: $p < k < n - p - \lfloor n\lambda_n \rfloor$. Then the upper and lower bound constants match in both cases and are given by formulas (3.6) and (3.8) for the case of $\lfloor n\lambda_n \rfloor \leq p$ and $p < \lfloor n\lambda_n \rfloor \leq n - 2p$, respectively. This follows immediately from the discussion in Section 6.3, in particular equations* (6.4) *and* (6.5)*.*

Remark 7. *It seems reasonable, that, in case of bumps of length smaller than p, we would need to analyze the type I error with some finer technique than just the union bound.*
On the other hand , we observe that if $n\lambda_n \to \infty$ and $\lambda_n \to 0$ as $n \to \infty$, then

$$\sup_{1 \leq k \leq \lfloor \lambda_n^{-1} \rfloor} \tilde{\sigma}_k \asymp n\lambda_n \left(\sum_{t=0}^{p} \phi_t \right)^2 \asymp \inf_{1 \leq k \leq \lfloor \lambda_n^{-1} \rfloor} \tilde{\sigma}_k,$$

in accordance with Theorem 4.

4 Simulations

In this Section we will perform numerical studies to examine the finite sample accuracy of the asymptotic upper bounds for the detection boundary. We focus on the situation that the noise ξ_n in (1.1) is generated by an AR(1) process, given by $\phi(z) = 1 - \rho z$ and $\theta(z) \equiv 1$ (in the notation of (3.1)), where $|\rho| < 1$. More precisely, the AR(1) process is given by the equation $Z_t - \rho Z_{t-1} = \zeta_t$, $t \in \mathbb{Z}$ where $\zeta_t \sim \mathcal{N}(0, 1)$ are i.i.d.. Note the slight difference to the setting considered in the introduction and Figure 1, as here the noise does not have standardized margins.

From Theorem 4 we obtain the detection boundary

$$\Delta_n \asymp \frac{\sqrt{2}}{1-\rho}\sqrt{\frac{-\log\lambda_n}{n\lambda_n}}. \tag{4.1}$$

In the following we fix the value of the detection rate $\sqrt{-\log\lambda_n/(n\lambda_n)}$ in (4.1) to be roughly $1/6$ and consider three different situations, namely small sample size ($\lambda_n = 0.1$, $n = 829$), medium sample size ($\lambda_n = 0.05$, $n = 2157$) and large sample size ($\lambda_n = 0.025$, $n = 5312$). Thus, the remaining free parameters are ρ and Δ_n, and the detection boundary (4.1) connects them by the asymptotic relation

$$\Delta_n \asymp \frac{\sqrt{2}}{1-\rho}\cdot\frac{1}{6} \approx \frac{0.236}{1-\rho}. \tag{4.2}$$

Let us now specify the investigated tests. For the (general) test from Section 2.2, the critical value $c_{\alpha,n}$ is only given implicitly, cf. (5.2). To simplify, in view of $c_{\alpha,n} = \sqrt{2\log\frac{2}{\alpha\lambda_n}}(1+o(1))$, we will therefore use

$$\Phi_n^{\mathrm{a}}(Y) := \begin{cases} 1 & \text{if } \sup\limits_{I\in\mathcal{I}(\lambda_n)} \frac{\left|\mathbf{1}_I^T Y\right|}{\sqrt{\mathbf{1}_I^T\Sigma_n\mathbf{1}_I}} > \sqrt{2\log\frac{2}{\alpha\lambda_n}}, \\ 0 & \text{else} \end{cases} \tag{4.3}$$

as an asymptotic version. Further we would like to investigate the maximum likelihood ratio test relying only on non-overlapping intervals $I_k = [(k-1)\lfloor n\lambda_n\rfloor + 1, k\lfloor n\lambda_n\rfloor)$ from Section 2.3 given by

$$\Phi_n^{\mathrm{d}}(Y) := \begin{cases} 1 & \text{if } \sup\limits_{1\le k\le\lfloor\frac{1}{\lambda_n}\rfloor} \frac{\left|\mathbf{1}_{I_k}^T\Sigma_n^{-1}Y\right|}{\sqrt{\mathbf{1}_{I_k}^T\Sigma_n^{-1}\mathbf{1}_{I_k}}} > \sqrt{2\log\frac{2}{\alpha\lambda_n}}, \\ 0 & \text{else.} \end{cases} \tag{4.4}$$

Note that the latter requires to scan only over $\lfloor 1/\lambda_n\rfloor$ intervals, whereas the former requires to scan over $\lfloor n(1-\lambda_n)\rfloor$ intervals. Consequently, the maximum likelihood ratio test from Section 2.3 can be computed faster by a factor of

$$\frac{n(1-\lambda_n)}{1/\lambda_n} = n\lambda_n(1-\lambda_n)$$

independent of Σ_n. For the three situations mentioned above this yields values of ≈ 74 in the small sample regime, ≈ 102 in the medium sample regime, and ≈ 129 in the large sample regime. However, our results from Theorem 1 and the discussion succeeding Theorem 2 imply, that the testing problems (1.3) and (2.3) are of the same difficulty in the sense that they both have the same separation rate.

In the following we examine the type I and type II errors $\bar{\alpha}(\Phi_*)$ and $\bar{\beta}(\Phi_*)$ with $* \in \{\mathrm{a},\mathrm{d}\}$ by 2000 simulation runs for $\alpha = 0.05$ with different choices of ρ, n, λ_n and Δ_n. The position $I \in \mathcal{I}(\lambda_n)$ is always drawn uniformly at random. Furthermore, we investigate the situation of 2 and 5 disjoint bumps within $[0,1]$.

The finite sample type I error of both Φ_n^{a} and Φ_n^{d} in all three sample size situations are shown in Figure 2 versus the correlation parameter $\rho \in \{-0.99, -0.98, ..., 0.99\}$. We find that Φ_n^{d} is somewhat conservative, which is clearly due to the usage of the union bound in deriving the critical value in (4.4), see the proof of Theorem 2. Opposed, Φ_n^{a} is conservative for $\rho > 0$, and liberal for $\rho < 0$. This is clearly due to the simplified critical value in (4.3), which is only asymptotically valid, and furthermore the employed asymptotics depend on ρ due to the result by Ibragimov and Linnik [36], see also Section 6.1 of Supplement A.2. However, it seems that already in the small sample size regime our asymptotic results provide a very good approximation for both tests.

Next we computed the finite sample type II error in all three sample size situations for $\rho \in \{-0.99, -0.98, ..., 0.99\}$ and $\Delta_n \in \{0.01, 0.02, ..., 0.5\}$. The corresponding results are shown in Figures 3–5. We also depict the contour line of equation (4.2) for a comparison and find a remarkably good agreement with the contour lines of the power function already in the small sample regime, which strongly supports the finite sample validity of our asymptotic theory. Finally, we conclude that detection becomes easier for a larger number of bumps.

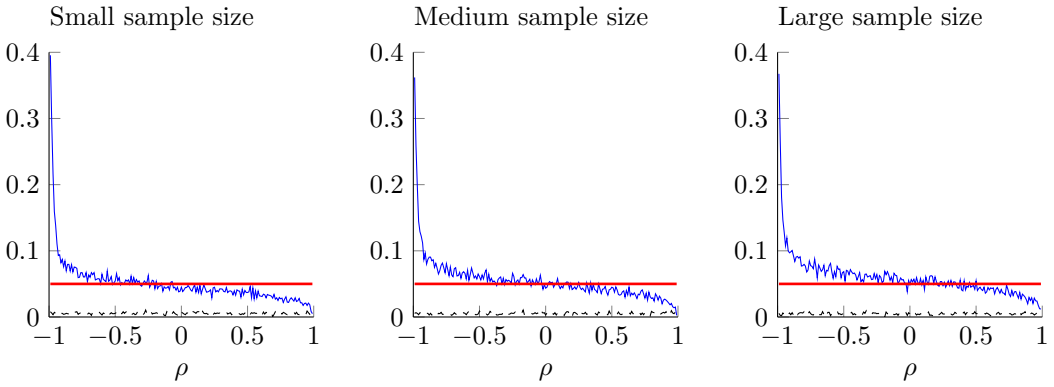


Figure 2: Type I error of the tests Φ_n^{a} (——) and Φ_n^{d} (- - -) for the AR(1) case vs. ρ (x-axis) simulated by 2000 Monte Carlo simulations with the nominal type I error $\alpha = 0.05$ (——) in three different situations: small sample size $\lambda_n = 0.1$, $n = 829$ (left), medium sample size $\lambda_n = 0.05$, $n = 2157$ (middle) and large sample size $\lambda_n = 0.025$, $n = 5312$ (right).

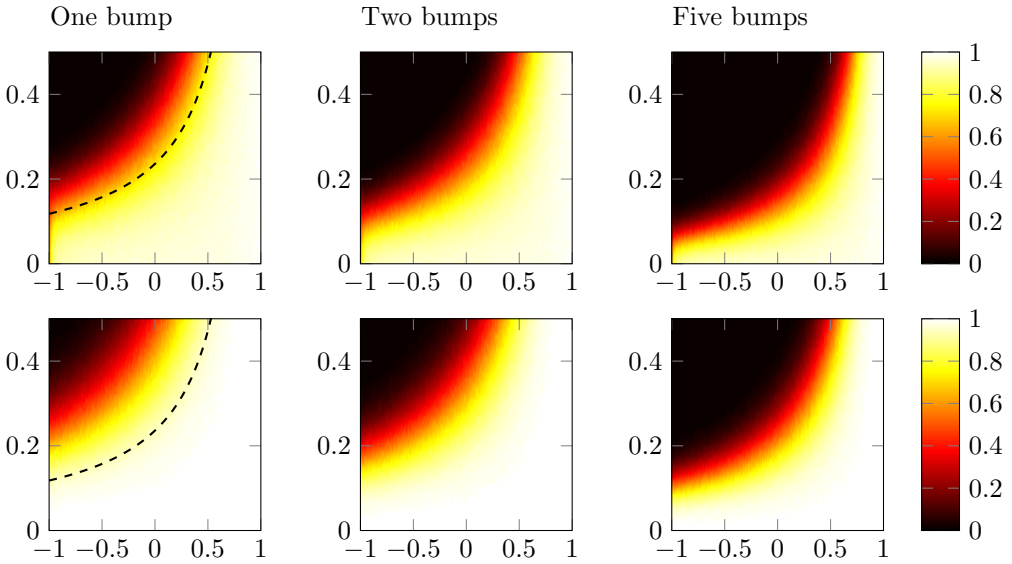


Figure 3: Type II error in the small sample regime $\lambda_n = 0.1$, $n = 829$ (top row: Φ_n^{a}; bottom row: Φ_n^{d}) for the AR(1) case for ρ (x-axis) vs. Δ_n (y-axis) with one bump (left column) together with the contour line of the detection boundary equation (4.2), two bumps (middle column) and five bumps (right column), each simulated by 2000 Monte Carlo simulations.

5 Proofs

Several useful results from various sources that we are going to use in our proofs can be found in Supplement A.2. The proof of Theorem 1 will then be split into three parts. We will provide asymptotic upper and lower bounds in subsections 5.1 and 5.1, respectively. The lower bound result will in fact hold for a wider class of covariance matrices than those allowed by Assumption 2. Finally, in subsection 5.1, this will be used to show that the upper and lower bound coincide asymptotically in the setting of Theorem 1, and this will yield the desired result. All remaining proofs can be found in Section 6 of Supplement A.2.

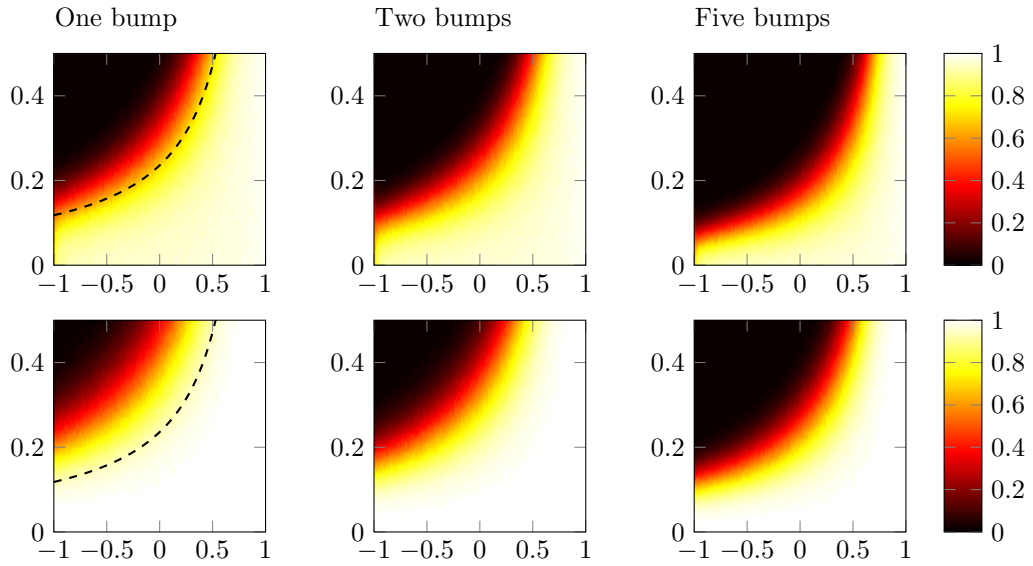


Figure 4: Type II error in the medium sample regime $\lambda_n = 0.05$, $n = 2157$ (top row: Φ_n^{a}; bottom row: Φ_n^{d}) for the AR(1) case for ρ (x-axis) vs. Δ_n (y-axis) with one bump (left column) together with the contour line of the detection boundary equation (4.2), two bumps (middle column) and five bumps (right column), each simulated by 2000 Monte Carlo simulations.

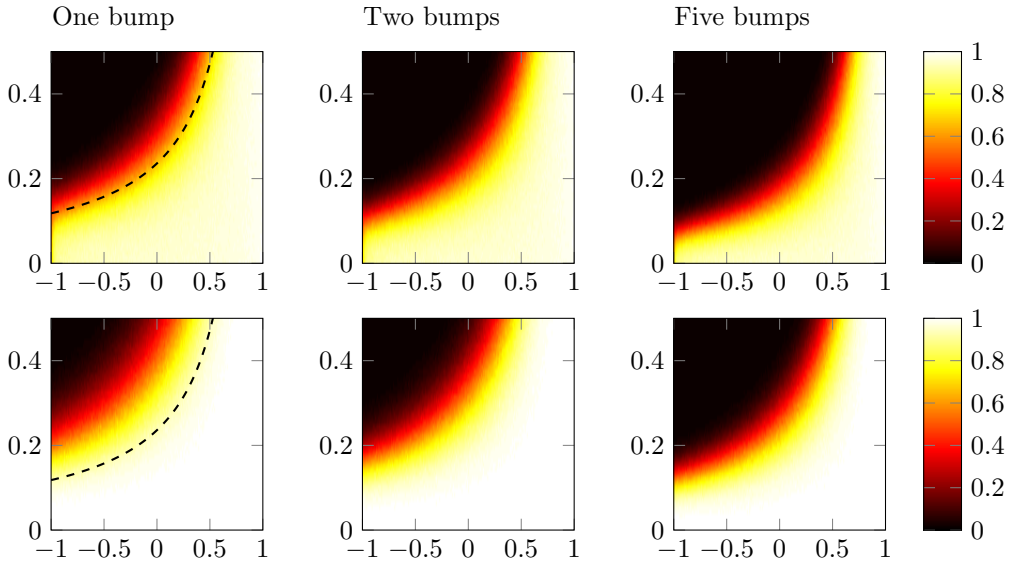


Figure 5: Type II error in the large sample regime $\lambda_n = 0.025$, $n = 5312$ (top row: Φ_n^{a}; bottom row: Φ_n^{d}) for the AR(1) case for ρ (x-axis) vs. Δ_n (y-axis) with one bump (left column) together with the contour line of the detection boundary equation (4.2), two bumps (middle column) and five bumps (right column), each simulated by 2000 Monte Carlo simulations.

5.1 Proofs for Section 2

Upper detection bound

For $I \in \mathcal{I}(\lambda_n)$ we define

$$\sigma_n(I) := \mathbf{1}_I^T \Sigma_n \mathbf{1}_I.$$

Lemma 8 (Upper detection bound)**.** *Fix $\alpha \in (0,1)$, consider the testing problem* (1.3) *and suppose that Assumption 1 and Assumption 2 hold. In addition, assume that the sequence $(\Sigma_n)_{n\in\mathbb{N}}$ of covariance matrices satisfies*

$$\Delta_n \inf_{I\in\mathcal{I}(\lambda_n)} \frac{\lfloor n\lambda_n \rfloor}{\sqrt{\sigma_n(I)}} \gtrsim (\sqrt{2} + \tilde{\varepsilon}_n)\sqrt{-\log \lambda_n}, \tag{5.1}$$

as $n \to \infty$, where $(\tilde{\varepsilon}_n)_{n\in\mathbb{N}}$ is a sequence of real numbers that satisfies $\tilde{\varepsilon}_n \to 0$ and $\tilde{\varepsilon}_n\sqrt{-\log \lambda_n} - \sqrt{\log\log n} \to \infty$ as $n \to \infty$.
Then the sequence of level α tests $(\Phi_n^{\mathrm{a}})_{n\in\mathbb{N}}$ as in (2.1) *with suitably chosen $c_{\alpha,n}$ satisfies $\bar{\alpha}(\Phi_n^{\mathrm{a}}) \le \alpha$ for all $n \in \mathbb{N}$ and $\limsup_{n\to\infty} \bar{\beta}(\Phi_n^{\mathrm{a}}) \le \alpha$.*

Proof. Define the test statistic

$$T_n(Y) = \sup_{I\in\mathcal{I}(\lambda_n)} \frac{\left|\mathbf{1}_I^T Y\right|}{\sqrt{\sigma_n(I)}},$$

and recall that $\Phi_n^{\mathrm{a}}(Y) := \mathbb{1}\{T_n(Y) > c_{\alpha,n}\}$, for some threshold $c_{\alpha,n}$ to be determined later.
We begin by noting that although the random process

$$\left(\frac{\left|\mathbf{1}_I^T Y\right|}{\sqrt{\sigma_n(I)}}\right)_{I\in\mathcal{I}(\lambda_n)}$$

has an infinite index set, it only takes finitely many different values. Thus, we can find a finite representative system $\mathcal{I}_{\mathrm{fin}}(\lambda_n) \subseteq \mathcal{I}(\lambda_n)$, such that for any $I \in \mathcal{I}(\lambda_n)$ there is $I' \in \mathcal{I}_{\mathrm{fin}}(\lambda_n)$, such that

$$\frac{\left|\mathbf{1}_I^T Y\right|}{\sqrt{\sigma_n(I)}} = \frac{\left|\mathbf{1}_{I'}^T Y\right|}{\sqrt{\sigma_n(I')}},$$

which implies that

$$T_n(Y) = \sup_{I\in\mathcal{I}_{\mathrm{fin}}(\lambda_n)} \frac{\left|\mathbf{1}_I^T Y\right|}{\sqrt{\sigma_n(I)}}$$

i.e. $T_n(Y)$ can be written as the supremum over the absolute values of a Gaussian process with a finite index set. Let M_n such that $\mathbb{E}_0 T_n(Y) \le M_n$. Then, for any $\lambda > 0$, it follows that

$$\begin{aligned}\mathbb{P}_0\left(T_n(Y) > \lambda + M_n\right) &\le \mathbb{P}_0\left(T_n(Y) - \mathbb{E}_0 T_n(Y) > \lambda\right) \\ &\le \mathbb{P}_0\left(\left|\sup_{I\in\mathcal{I}_{\mathrm{fin}}(\lambda_n)} \frac{\left|\mathbf{1}_I^T Y\right|}{\sqrt{\sigma_n(I)}} - \mathbb{E}_0 \sup_{I\in\mathcal{I}_{\mathrm{fin}}(\lambda_n)} \frac{\left|\mathbf{1}_I^T Y\right|}{\sqrt{\sigma_n(I)}}\right| > \lambda\right) \le 2e^{-\frac{\lambda^2}{2}},\end{aligned}$$

where the last inequality follows the results of Talagrand [68] and can be found in Theorem 2.1.20 of [28]. Thus, if we let

$$c_{\alpha,n} = \sqrt{2\log\frac{2}{\alpha}} + M_n,$$

Φ_n^{a} has level α for any n.
In order to find a suitable bound M_n we consider an even coarser finite subset of $\mathcal{I}(\lambda_n)$. Let

$$\mathcal{C}_n = \left\{\left[\frac{k}{n}, \frac{k}{n} + \lambda_n\right) : 1 \le k \le \lfloor n(1-\lambda_n)\rfloor\right\} \subseteq \mathcal{I}(\lambda_n).$$

Clearly, $\#\mathcal{C}_n = \lfloor n(1-\lambda_n)\rfloor \leq n < \infty$. For any $I \in \mathcal{I}(\lambda_n)$ there is $I' \in \mathcal{C}_n$, such that 1_I differs from $1_{I'}$ in at most one entry. Thus, it is easy to see that

$$T_n(Y) \leq \sup_{I\in\mathcal{C}_n} \frac{\left|\mathbf{1}_I^T Y\right|}{\sqrt{\sigma_n(I)}} + \frac{\sup_{1\leq i\leq n}|Y_{i,n}|}{\inf_{I\in\mathcal{I}(\lambda_n)}\sqrt{\sigma_n(I)}},$$

Thus, we can set

$$M_n = \tilde{M}_n + \kappa_n,$$

where

$$\tilde{M}_n = \mathbb{E}_0 \sup_{I\in\mathcal{C}_n} \frac{\left|\mathbf{1}_I^T Y\right|}{\sqrt{\sigma_n(I)}},$$

and

$$\kappa_n = \mathbb{E}_0 \frac{\sup_{1\leq i\leq n}|Y_{i,n}|}{\inf_{I\in\mathcal{I}(\lambda_n)}\sqrt{\sigma_n(I)}}.$$

The latter term is easy to handle: We have

$$\mathbb{E}_0 \sup_{1\leq i\leq n} |Y_{i,n}| \leq \sqrt{2f_0\log(2n)},$$

since $Y_{i,n}$ has variance f_0 for any n and $1 \leq i \leq n$, and

$$\inf_{I\in\mathcal{I}(\lambda_n)} \sigma_n(I) = n\lambda_n f(0)(1+o(1))$$

by Theorem 18.2.1 of Ibragimov and Linnik [36]. Thus,

$$\kappa_n = O\left(\sqrt{\frac{\log n}{n\lambda_n}}\right),$$

and thus, $\kappa_n \to 0$ by Assumption 1. The next part of the proof will be devoted to computing $\tilde{M}_n$. Note that under H_0, we have

$$\frac{\mathbf{1}_I^T Y}{\sqrt{\sigma_n(I)}} \sim \mathcal{N}(0,1),$$

for any $I \in \mathcal{C}_n$. For any $I, I' \in \mathcal{C}_n$, we have

$$\left|\frac{\mathbf{1}_I^T Y}{\sqrt{\sigma_n(I)}} - \frac{\mathbf{1}_{I'}^T Y}{\sqrt{\sigma_n(I')}}\right| = \left|\left(\frac{\mathbf{1}_I^T}{\sqrt{\sigma_n(I)}} - \frac{\mathbf{1}_{I'}^T}{\sqrt{\sigma_n(I')}}\right) Y\right| = \left(2 - 2\frac{\mathbf{1}_I^T\Sigma_n\mathbf{1}_{I'}}{\sqrt{\sigma_n(I)\sigma_n(I')}}\right)^{\frac{1}{2}} |Z_{I,I'}|,$$

for some random variable $Z_{I,I'} \sim \mathcal{N}(0,1)$. Note that the system $\{Z_{I,I'} : I, I' \in \mathcal{C}_n\}$ is not necessarily independent. Let

$$d_n(I,I') := \left(2 - 2\frac{\mathbf{1}_I^T\Sigma_n\mathbf{1}_{I'}}{\sigma_n(I)}\right)^{\frac{1}{2}}.$$

Since Σ_n is a Toeplitz matrix, it follows that $\sigma_n(I) = \sigma_n(I')$ for any $I, I' \in \mathcal{C}_n$, and thus, $d_n(I,I') = d_n(I',I)$. Since Σ_n is also positive definite, it is then easy to see that d_n is a metric on $\mathcal{C}_n$.
Now let $\mathcal{E}_n \subseteq \mathcal{C}_n$ be an η_n-net for $(\mathcal{C}_n, d_n)$, i.e. for any $I \in \mathcal{D}_n$ there is $J \in \mathcal{E}_n$, such that

$$d_n(I,J) \leq \eta_n.$$

For any $I \in \mathcal{C}_n$ and $J \in \mathcal{E}_n$ we have

$$\frac{\left|\mathbf{1}_I^T Y\right|}{\sqrt{\sigma_n(I)}} \leq \left|\frac{\mathbf{1}_I^T Y}{\sqrt{\sigma_n(I)}} - \frac{\mathbf{1}_J^T Y}{\sqrt{\sigma_n(J)}}\right| + \frac{\left|\mathbf{1}_J^T Y\right|}{\sqrt{\sigma_n(J)}},$$

and thus,

$$\begin{aligned}\sup_{I\in\mathcal{C}_n}\frac{|\mathbf{1}_I^TY|}{\sqrt{\sigma_n(I)}} &\le \sup_{I\in\mathcal{C}_n}\inf_{J\in\mathcal{E}_n}\left|\frac{\mathbf{1}_I^TY}{\sqrt{\sigma_n(I)}}-\frac{\mathbf{1}_J^TY}{\sqrt{\sigma_n(J)}}\right|+\sup_{J\in\mathcal{E}_n}\frac{|\mathbf{1}_J^TY|}{\sqrt{\sigma_n(J)}}\\ &= \sup_{I\in\mathcal{C}_n}\inf_{J\in\mathcal{E}_n} d_n(I,J)|Z_{I,J}|+\sup_{J\in\mathcal{E}_n}\frac{|\mathbf{1}_J^TY|}{\sqrt{\sigma_n(J)}}\\ &\le \eta_n\sup_{I\in\mathcal{C}_n}\inf_{J\in\mathcal{E}_n}|Z_{I,J}|+\sup_{J\in\mathcal{E}_n}\frac{|\mathbf{1}_J^TY|}{\sqrt{\sigma_n(J)}}.\end{aligned}$$

It follows that

$$\begin{aligned}\tilde{M}_n &\le \eta_n\sqrt{2\log(2n)}+\sqrt{2\log\left(2N\left(\mathcal{C}_n,d_n,\eta_n\right)\right)}\\ &\le \eta_n\sqrt{2\log n}+\sqrt{2\log N\left(\mathcal{C}_n,d_n,\eta_n\right)}+(1+\eta_n)\sqrt{2\log 2},\end{aligned}$$

where $N\left(\mathcal{C}_n,d_n,\eta_n\right)$ is the η_n-covering number of $(\mathcal{C}_n,d_n)$. Now let $I,I'\in\mathcal{C}_n$, $I\neq I'$, with $d_H(I,I')\le\frac{\lambda_n}{\log n}$, where d_H denotes the Hausdorff metric on the set of subintervals of $[0,1]$ (with respect to the euclidean distance on $[0,1]$), i.e. $d_H(I,I')=|\inf I-\inf I'|$. In addition, we assume that $\inf I<\inf I'$ without loss of generality. Then

$$\begin{aligned}\mathbf{1}_I^T\Sigma_n\mathbf{1}_{I'} &= (\mathbf{1}_{I\cap I'}^T+\mathbf{1}_{I\setminus I'}^T)\Sigma_n(\mathbf{1}_{I\cap I'}+\mathbf{1}_{I'\setminus I})\\ &= \mathbf{1}_{I\cap I'}^T\Sigma_n\mathbf{1}_{I\cap I'}+\mathbf{1}_{I\setminus I'}^T\Sigma_n\mathbf{1}_{I\cap I'}+\mathbf{1}_{I\cap I'}^T\Sigma_n\mathbf{1}_{I'\setminus I}+\mathbf{1}_{I\setminus I'}^T\Sigma_n\mathbf{1}_{I'\setminus I}.\end{aligned}$$

Due to Σ_n being symmetric and Toeplitz, we have $\mathbf{1}_{I\cap I'}^T\Sigma_n\mathbf{1}_{I'\setminus I}=\mathbf{1}_{I\cap I'}^T\Sigma_n\mathbf{1}_{I\setminus I'}$, and thus,

$$\mathbf{1}_I^T\Sigma_n\mathbf{1}_{I'}=\mathbf{1}_I^T\Sigma_n\mathbf{1}_I-\mathbf{1}_{I\setminus I'}^T\Sigma_n\mathbf{1}_{I\setminus I'}+\mathbf{1}_{I\setminus I'}^T\Sigma_n\mathbf{1}_{I'\setminus I}.$$

It follows that

$$\begin{aligned}d_n^2(I,I') &= 2-2\left(\mathbf{1}_I^T\Sigma_n\mathbf{1}_I\right)^{-1}\left[\mathbf{1}_I^T\Sigma_n\mathbf{1}_I-\mathbf{1}_{I\setminus I'}^T\Sigma_n\mathbf{1}_{I\setminus I'}+\mathbf{1}_{I\setminus I'}^T\Sigma_n\mathbf{1}_{I'\setminus I}\right]\\ &= 2\left(\mathbf{1}_I^T\Sigma_n\mathbf{1}_I\right)^{-1}\left[\mathbf{1}_{I\setminus I'}^T\Sigma_n\mathbf{1}_{I\setminus I'}-\mathbf{1}_{I\setminus I'}^T\Sigma_n\mathbf{1}_{I'\setminus I}\right].\end{aligned}$$

Since $\mathbf{1}_{I\setminus I'}^T\Sigma_n\mathbf{1}_{I'\setminus I}$ is the sum over a submatrix with $r_n=n|\inf I-\inf I'|$ rows, and its lower left entry is $f_{\lfloor n\lambda_n\rfloor-1-r_n}$, we find the trivial bound

$$\left|\mathbf{1}_{I\setminus I'}^T\Sigma_n\mathbf{1}_{I'\setminus I}\right|\le\frac{n\lambda_n}{\log n}\sum_{h=\lfloor n\lambda_n(1-1/\log n)\rfloor-1}^{\infty}M(1+|h|)^{-1-\kappa}=o\left(\frac{n\lambda_n}{\log n}\right),$$

as $n\to\infty$. We use Theorem 18.2.1 of Ibragimov and Linnik [36] (see also Section 6.1 of Supplement A.2) to find $\mathbf{1}_I^T\Sigma_n\mathbf{1}_I=n\lambda_nf(0)\Big(1+o(1)\Big)$ and

$$\mathbf{1}_{I\setminus I'}^T\Sigma_n\mathbf{1}_{I\setminus I'}\le f(0)\frac{n\lambda_n}{\log n}(1+o(1)),$$

as $n\to\infty$. This yields

$$d_n(I,I')\le\sqrt{\frac{2}{\log n}}+\zeta_n,$$

where $\zeta_n=o\left((\log n)^{-\frac{1}{2}}\right)$. This implies that for any $I\in\mathcal{C}_n$ and for large enough n we have the inclusion

$$\left\{I'\in\mathcal{C}_n:d_H(I,I')\le\frac{\lambda_n}{\log n}\right\}\subseteq\left\{I'\in\mathcal{C}_n:d_n(I,I')\le\sqrt{\frac{2}{\log n}}+\zeta_n\right\}.$$

Thus, if we choose $\eta_n = \sqrt{\frac{2}{\log n}} + \zeta_n$, this yields the bound

$$N\left(\mathcal{C}_n, d_n, \eta_n\right) \leq \frac{\log n}{2\lambda_n},$$

and consequently,

$$\tilde{M}_n \leq 2 + \zeta_n\sqrt{2\log n} + \sqrt{2\log\frac{\log n}{2\lambda_n}} + \left(1 + \zeta_n + \sqrt{\frac{2}{\log n}}\right)\sqrt{2\log 2}.$$

Thus, if we choose

$$c_{\alpha,n} = 2 + \zeta_n\sqrt{2\log n} + \sqrt{2\log\frac{2}{\alpha}} + \sqrt{2\log\frac{\log n}{2\lambda_n}} + \kappa_n + \left(1 + \zeta_n + \sqrt{\frac{2}{\log n}}\right)\sqrt{2\log 2}, \quad (5.2)$$

the test Φ_n^{a} will have level α for all $n \in \mathbb{N}$. Note that $\zeta_n\sqrt{2\log n} = o(1)$ as $n \to \infty$. Concerning the type II error of the test Φ_n^{a}, recall that, under H_1, i.e. if $Y \sim \mathcal{N}\left(\delta \mathbf{1}_I, \Sigma_n\right)$ for some δ with $|\delta| > \Delta_n$, and $I \in \mathcal{I}(\lambda_n)$, we have

$$\frac{\mathbf{1}_{I'}^T Y}{\sqrt{\sigma_n(I')}} \sim \mathcal{N}\left(\frac{\delta \mathbf{1}_{I'}^T \mathbf{1}_I}{\sqrt{\sigma_n(I')}}, 1\right).$$

for all $I' \in \mathcal{I}(\lambda_n)$. For n large enough, it follows from plugging in (5.1) and (5.2), that

$$\begin{aligned}
\bar{\beta}(\Phi_n^{\mathrm{a}}) &= \sup_{I\in\mathcal{I}(\lambda_n)} \sup_{|\delta|\geq\Delta_n} \mathbb{P}_{I,\delta}\left[\Phi_n^{\mathrm{a}}(Y) = 0\right] \\
&= \sup_{I\in\mathcal{I}(\lambda_n)} \sup_{|\delta|\geq\Delta_n} \mathbb{P}\left[\sup_{I'\in\mathcal{I}(\lambda_n)} \left|Z_{I'} + \frac{\delta \mathbf{1}_{I'}^T \mathbf{1}_I}{\sqrt{\sigma_n(I')}}\right| \leq c_{\alpha,n}\right] \\
&\leq \sup_{I\in\mathcal{I}(\lambda_n)} \sup_{|\delta|\geq\Delta_n} \mathbb{P}\left[\left|Z_I + \frac{\delta \mathbf{1}_I^T \mathbf{1}_I}{\sqrt{\sigma_n(I)}}\right| \leq c_{\alpha,n}\right] \\
&\leq \sup_{I\in\mathcal{I}(\lambda_n)} \sup_{|\delta|\geq\Delta_n} \mathbb{P}\left[|\delta|\frac{\mathbf{1}_I^T \mathbf{1}_I}{\sqrt{\sigma_n(I)}} - |Z_I| \leq c_{\alpha,n}\right] \\
&\leq \mathbb{P}\left[|Z| > \Delta_n \inf_{I\in\mathcal{I}(\lambda_n)} \frac{\mathbf{1}_I^T \mathbf{1}_I}{\sqrt{\sigma_n(I)}} - c_{\alpha,n}\right],
\end{aligned}$$

where $(Z_I)_{I\in\mathcal{I}(\lambda_n)}$ and Z are (not necessarily independent) standard Gaussian random variables. Plugging in (5.1), we have

$$\begin{aligned}
&\Delta_n \inf_{I\in\mathcal{I}(\lambda_n)} \frac{\mathbf{1}_I^T \mathbf{1}_I}{\sqrt{\sigma_n(I)}} - c_{\alpha,n} \\
&\geq \tilde{\varepsilon}_n\sqrt{\log\frac{1}{\lambda_n}} - 2 - \zeta_n\sqrt{2\log n} - \sqrt{2\log\frac{2}{\alpha}} - \sqrt{2\log\frac{\log n}{2}} - \kappa_n - \left(1 + \zeta_n + \sqrt{\frac{2}{\log n}}\right)\sqrt{2\log 2},
\end{aligned}$$

for n large enough. Since $\tilde{\varepsilon}_n\sqrt{-\log\lambda_n} - \sqrt{\log\log n} \to \infty$ by assumption and $\kappa_n = o(1)$ and $\zeta_n\sqrt{2\log n} = o(1)$ as $n \to \infty$, it follows that the right-hand side diverges to ∞. This finishes the proof. □

Lower detection bound

We start by giving some technicalities on LR-statistics required throughout the paper at several places. As λ_n and Σ_n are known, the likelihood ratio $L_{\delta,I} = L_{\delta,I}(Y)$ between the distributions of Y under H_0 and $H_{\delta,I}^n$ is given by

$$L_{I,\delta} = \exp\left[\delta \mathbf{1}_I^T \Sigma_n^{-1} Y - \frac{1}{2}\delta^2 \mathbf{1}_I^T \Sigma_n^{-1} \mathbf{1}_I\right].$$

Note that, under H_0, the likelihood ratio $L_{\delta,I}$ follows a log-normal distribution, i.e.

$$\log L_{I,\delta} = \delta \mathbf{1}_I^T \Sigma_n^{-1} Y - \frac{1}{2}\delta^2 \mathbf{1}_I^T \Sigma_n^{-1} \mathbf{1}_I \overset{H_0}{\sim} \mathcal{N}_1\left(-\frac{1}{2}\delta^2 \tilde{\sigma}_n(I), \delta^2 \tilde{\sigma}_n(I)\right),$$

where

$$\tilde{\sigma}_n(I) := \mathbf{1}_I^T \Sigma_n^{-1} \mathbf{1}_I.$$

Note that, under H_0, for $I, I' \in \mathcal{C}_n$ and $\delta \in \mathbb{R}$, we have $\mathbb{E}L_{I,\delta} = 1$, $\operatorname{Var} L_{I,\delta} = \exp\left(\delta^2 \tilde{\sigma}_n(I)\right) - 1$ and $\operatorname{Cov}(L_{I,\delta}, L_{I',\delta}) = \exp\left(\delta^2 \mathbf{1}_I^T \Sigma_n^{-1} \mathbf{1}_{I'}\right) - 1$. Finally, let

$$\mathcal{I}^0 := \left\{[(k-1)\lambda_n, k\lambda_n) : 1 \le k \le \lfloor \lambda_n^{-1} \rfloor\right\} \subseteq \mathcal{I}(\lambda_n)$$

be a system of non-overlapping intervals of length λ_n as defined in (2.2).

Lemma 9 (Lower detection bound)**.** *Fix $\alpha \in (0,1)$, and suppose that (1.5) holds. Let $(\Sigma_n)_{n\in\mathbb{N}}$ be a sequence of covariance matrices, such that*

$$\Delta_n \sup_{I \in \mathcal{I}(\lambda_n)} \sqrt{\tilde{\sigma}_n(I)} \precsim \left(\sqrt{2} - \varepsilon_n\right) \sqrt{-\log \lambda_n}, \tag{5.3}$$

where $(\varepsilon_n)_{n\in\mathbb{N}}$ is a sequence that satisfies $\varepsilon_n \to 0$ and $\varepsilon_n \sqrt{-\log \lambda_n} \to \infty$ as $n \to \infty$. In addition, assume that for some $m \in \mathbb{N}_0$

$$\lim_{n\to\infty} \frac{1}{\lfloor \lambda_n^{-1} \rfloor^2} \sum_{\substack{I, I' \in \mathcal{I}^0 \\ n|\inf I - \inf I'| > m}} \operatorname{Cov}\left(L_{I,\Delta_n}, L_{I',\Delta_n}\right) = 0, \tag{5.4}$$

as $n \to \infty$.
Then any sequence of tests $(\Phi_n)_{n\in\mathbb{N}}$ with $\limsup_{n\to\infty} \bar{\alpha}(\Phi_n) \le \alpha$ will obey $\limsup_{n\to\infty} \bar{\beta}(\Phi_n) \ge 1 - \alpha$, i.e. the bump is asymptotically undetectable.

Proof. We employ the same strategy as in the proof of Theorem 3.1(a) of Dümbgen and Spokoiny [19]. We bound the type II error of any given test by an expression that does not depend on the test anymore, and then employ an appropriate L^1-law of large numbers for dependent arrays of random variables.

For any sequence of tests Φ_n with asymptotic level α under H_0 we have

$$\begin{aligned}
\bar{\beta}(\Phi_n) &= \sup_{I \in \mathcal{I}(\lambda_n)} \sup_{|\delta| \ge \Delta_n} \mathbb{E}_{I,\delta}\left[1 - \Phi_n(Y)\right] \ge \sup_{I \in \mathcal{I}^0} \sup_{|\delta| \ge \Delta_n} \mathbb{E}_{I,\delta}\left[1 - \Phi_n(Y)\right] \\
&\ge \frac{1}{\lfloor \lambda_n^{-1} \rfloor} \sum_{I \in \mathcal{I}^0} \sup_{|\delta| \ge \Delta_n} \mathbb{E}_{I,\delta}\left[1 - \Phi_n(Y)\right] \ge 1 - \frac{1}{\lfloor \lambda_n^{-1} \rfloor} \sum_{I \in \mathcal{I}^0} \mathbb{E}_{I,\Delta_n}\left[\Phi_n(Y)\right] \\
&\ge 1 - \frac{1}{\lfloor \lambda_n^{-1} \rfloor} \sum_{I \in \mathcal{I}^0} \mathbb{E}_0\left[\Phi_n(Y) \frac{d\mathbb{P}_{I,\Delta_n}}{d\mathbb{P}_0} - \Phi_n(Y)\right] - \alpha + o(1) \\
&= 1 - \mathbb{E}_0\left[\left(\frac{1}{\lfloor \lambda_n^{-1} \rfloor} \sum_{I \in \mathcal{I}^0} L_{I,\Delta_n} - 1\right) \Phi_n(Y)\right] - \alpha + o(1) \\
&\ge 1 - \alpha - \mathbb{E}_0\left|\frac{1}{\lfloor \lambda_n^{-1} \rfloor} \sum_{I \in \mathcal{I}^0} L_{I,\Delta_n} - 1\right| + o(1).
\end{aligned}$$

Next, we show that the array $\left\{L_{\Delta_n,I} : I \in \mathcal{I}^0, n \in \mathbb{N}\right\}$ is *h-integrable with exponent 1* (see Definition 12 in Supplement A.2 or Definition 1.5 in Sung, Lisawadi and Volodin [67]), i.e. we show that

$$\sup_{n\in\mathbb{N}} \frac{1}{\lfloor \lambda_n^{-1} \rfloor} \sum_{I \in \mathcal{I}^0} \mathbb{E}_0\left[|L_{I,\Delta_n}|\right] < \infty, \quad \text{and} \quad \lim_{n\to\infty} \frac{1}{\lfloor \lambda_n^{-1} \rfloor} \sum_{I \in \mathcal{I}^0} \mathbb{E}_0\left[|L_{I,\Delta_n}| \mathbf{1}\left\{|L_{I,\Delta_n}| > h(n)\right\}\right] = 0, \tag{5.5}$$

where $h(n) = \lfloor \lambda_n^{-1} \rfloor^{\frac{1}{2}(1+\varepsilon_n)\left(\sqrt{2}-\varepsilon_n\right)^2}$. Since $\mathbb{E}_0|L_{I,\Delta_n}| = 1$ for all $n \in \mathbb{N}$ and $I \in \mathcal{I}^0$, the first condition is satisfied.

Further, if n large is enough, we have

$$\begin{aligned}
\frac{1}{\lfloor \lambda_n^{-1} \rfloor} \sum_{I \in \mathcal{I}^0} \mathbb{E}_0 \left[L_{I,\Delta_n} \mathbf{1}\left\{ L_{I,\Delta_n} > h(n) \right\} \right] &\leq \sup_{I \in \mathcal{I}^0} \mathbb{E}_0 \left[L_{I,\Delta_n} \mathbf{1}\left\{ L_{I,\Delta_n} > h(n) \right\} \right] \\
&= \sup_{I \in \mathcal{I}^0} \mathbb{P}\left(Z \leq \frac{\frac{1}{2}\Delta_n^2 \tilde{\sigma}_n(I) - \log h(n)}{\Delta_n \sqrt{\tilde{\sigma}_n(I)}} \right) \\
&\leq \mathbb{P}\left(Z \leq \sup_{I \in \mathcal{I}^0} \frac{1}{2}\Delta_n \sqrt{\tilde{\sigma}_n(I)} - \frac{\log h(n)}{\sup_{I \in \mathcal{I}^0} \Delta_n \sqrt{\tilde{\sigma}_n(I)}} \right) \\
&\overset{\text{(a)}}{\leq} \mathbb{P}\left(Z \leq -\varepsilon_n(\sqrt{2} - \varepsilon_n)\sqrt{-\log \lambda_n} \right),
\end{aligned}$$

where Z is a standard Gaussian random variable. The inequality (a) follows immediately from (5.3) and the definition of $h(n)$. The claim follows from the assumption that $\lim_{n\to\infty} \varepsilon_n \sqrt{-\log \lambda_n} = \infty$ as $n \to \infty$.

Then, given that (5.4) and (5.5) hold, it follows from an L^1-law of large numbers for dependent arrays (see Theorem 13 in Supplement A.2 or Theorem 3.2 of Wang and Hu [71]), that

$$\mathbb{E}_0 \left| \frac{1}{\lfloor \lambda_n^{-1} \rfloor} \sum_{I \in \mathcal{I}^0} L_{\Delta_n, I} - 1 \right| \to 0, \tag{5.6}$$

as $n \to \infty$, which finishes the proof. □

Proof of Theorem 1

In the setting described in Theorem 1 the noise vector ξ_n in model (1.1) is given by n consecutive realizations of a stationary centered Gaussian process with the square summable autocovariance function $\gamma(h)$, $h \in \mathbb{Z}$ and the spectral density f. We suppose that Assumption 2 is satisfied, i.e. the autocovariance of ξ_n has a polynomial decay. In terms of Σ_n, this means

$$|\Sigma_n(i,j)| \leq C(1 + |i-j|)^{-(1+\kappa)},$$

for $1 \leq i, j \leq n$ and some constants $C > 0$ and $\kappa > 0$.

In order to apply Lemma 9 in such a setting, first, we need to examine the asymptotic behavior of the coefficients $\tilde{\sigma}_n(I)$, and second, we need to verify that condition (5.4) is satisfied under the lower detection boundary condition (5.3) and Assumption 2.

For the setting of Theorem 1, we will do the former in Lemma 10 and the latter in Lemma 11.

Lemma 10. *If Assumption 2 holds, then for any $I \in \mathcal{I}(\lambda_n)$, it follows that*

$$\tilde{\sigma}_n(I) = \frac{n\lambda_n}{f(0)}(1 + o(1)),$$

as $n \to \infty$.

The proof of Lemma 10 is very similar to the proof of Proposition C.1 in Keshavarz, Scott and Nguyen [45], and can be found in Supplement A.2.

Lemma 11. *If Assumption 2 holds, and given that*

$$\Delta_n \sup_{I \in \mathcal{I}(\lambda_n)} \sqrt{\tilde{\sigma}_n(I)} \precsim (\sqrt{2} - \varepsilon_n)\sqrt{-\log \lambda_n} \tag{5.7}$$

for a sequence $(\varepsilon_n)_{n\in\mathbb{N}}$ *satisfying* $\varepsilon_n \to 0$ *and* $\varepsilon_n\sqrt{-\log\lambda_n} \to \infty$ *as* $n \to \infty$*, then condition* (5.4) *holds with* $m = 1$*, i.e.*

$$\lim_{n\to\infty} \lambda_n^2 \sum_{\substack{I,I'\in\mathcal{I}^0 \\ n|\inf I - \inf I'|>1}} \exp\left(\Delta_n^2 \mathbf{1}_I^T \Sigma_n^{-1} \mathbf{1}_{I'}\right) - 1 = 0,$$

Proof. For $I, I' \in \mathcal{I}^0$ with $n|\inf I - \inf I'| > 1$. Write

$$\begin{aligned}\exp\left(\Delta_n^2 \mathbf{1}_I^T \Sigma_n^{-1} \mathbf{1}_{I'}\right) - 1 &= \sum_{p=1}^{\infty} \frac{1}{p!} \left[\Delta_n^2 \mathbf{1}_I^T \Sigma_n^{-1} \mathbf{1}_{I'}\right]^p \\ &= \sum_{p=1}^{\infty} \frac{1}{p!} \left[\frac{1}{2}\Delta_n^2 \sqrt{\tilde{\sigma}_n(I)\tilde{\sigma}_n(I')}\right]^p \left[2\frac{\mathbf{1}_I^T \Sigma_n^{-1} \mathbf{1}_{I'}}{\sqrt{\tilde{\sigma}_n(I)\tilde{\sigma}_n(I')}}\right]^p.\end{aligned}$$

If $n\lambda_n$ is an integer, the latter term $\mathbf{1}_I^T \Sigma_n^{-1} \mathbf{1}_{I'}$ is the sum over a square submatrix of Σ_n^{-1}, and if $n\lambda_n$ is not an integer, then the number of non-zero entries of $\mathbf{1}_I$ and $\mathbf{1}_{I'}$ cannot differ by more than 1. From Lemma A.1 of [31] (see also Section 6.1 in Supplement A.2, it trivially follows that

$$\begin{aligned}\left|\mathbf{1}_I^T \Sigma_n^{-1} \mathbf{1}_{I'}\right| &\leq C' \lceil n\lambda_n \rceil \sum_{t=1}^{\lceil n\lambda_n \rceil} \left(n|\inf I - \inf I'| \lfloor n\lambda_n \rfloor + t\right)^{-(1+\kappa)} \\ &\leq C' \lceil n\lambda_n \rceil \sum_{t=1}^{\lceil n\lambda_n \rceil} \left(\lfloor n\lambda_n \rfloor\right)^{-(1+\kappa)} = o\left(n\lambda_n\right).\end{aligned}$$

From Lemma 10, we know that $\sqrt{\tilde{\sigma}_n(I)\tilde{\sigma}_n(I')} = \frac{n\lambda_n}{f(0)}(1 + o(1))$ as $n \to \infty$, and thus, it follows that $\sqrt{\tilde{\sigma}_n(I)\tilde{\sigma}_n(I')}^{-1} \mathbf{1}_I^T \Sigma_n^{-1} \mathbf{1}_{I'} \to 0$ as $n \to \infty$. Hence, for n large enough, we have

$$\begin{aligned}&\left|\sum_{p=1}^{\infty} \frac{1}{p!} \left[\frac{1}{2}\Delta_n^2 \sqrt{\tilde{\sigma}_n(I)\tilde{\sigma}_n(I')}\right]^p \left[2\frac{\mathbf{1}_I^T \Sigma_n^{-1} \mathbf{1}_{I'}}{\sqrt{\tilde{\sigma}_n(I)\tilde{\sigma}_n(I')}}\right]^p\right| \\ &\qquad\qquad \leq 2\frac{\left|\mathbf{1}_I^T \Sigma_n^{-1} \mathbf{1}_{I'}\right|}{\sqrt{\tilde{\sigma}_n(I)\tilde{\sigma}_n(I')}} \exp\left[\frac{1}{2}\Delta_n^2 \sqrt{\tilde{\sigma}_n(I)\tilde{\sigma}_n(I')}\right]. \quad (5.8)\end{aligned}$$

Note that from the lower detection boundary condition (5.7) it immediately follows that

$$\exp\left[\frac{1}{2}\Delta_n^2 \sqrt{\tilde{\sigma}_n(I)\tilde{\sigma}_n(I')}\right] \leq \lambda_n^{-\frac{1}{2}(\sqrt{2}-\varepsilon_n)^2} \leq \lambda_n^{-1} \quad (5.9)$$

for n large enough. Applying Lemma A.1 of [31] again, it follows that $|\Sigma_n^{-1}(i,j)| \leq C(1 + |i - j|)^{-(1+\kappa)}$ for some $C > 0$. Let Φ_n be the $n \times n$-matrix with entries $\Phi_n(i,j) = C(1 + |i-j|)^{-(1+\kappa)}$, and let $\Phi(\nu) = \sum_{h=-\infty}^{\infty} C(1 + |i - j|)^{-(1+\kappa)} e^{-2\pi i h\nu}$. Then

$$\sum_{\substack{I,I'\in\mathcal{I}^0 \\ n|\inf I - \inf I'|>1}} |\mathbf{1}_I^T \Sigma_n^{-1} \mathbf{1}_{I'}| \leq \sum_{\substack{I,I'\in\mathcal{I}^0 \\ I\neq I'}} \mathbf{1}_I^T \Phi_n \mathbf{1}_{I'} \leq \sum_{i,j=1}^{n} \Phi_n(i,j) - \sum_{I\in\mathcal{I}^0} \mathbf{1}_I^T \Phi_n \mathbf{1}_I \overset{(a)}{=} o(n), \quad (5.10)$$

where (a) follows from Theorem 18.2.1 of Ibragimov and Linnik [36], since it yields that $\sum_{i,j=1}^{n} \Phi_n(i,j) = n\Phi(0) + o(n)$ and $\mathbf{1}_I^T \Phi_n \mathbf{1}_I = n\lambda_n \Phi(0) + o\left(n\lambda_n\right)$ for any $I \in \mathcal{I}^0$.

Finally, combining (5.8), (5.9) and (5.10), and once again using that $\sqrt{\tilde{\sigma}_n(I)\tilde{\sigma}_n(I')} = \frac{n\lambda_n}{f(0)}(1+o(1))$ as $n \to \infty$, we find

$$\sum_{\substack{I,I'\in\mathcal{I}^0 \\ n|\inf I - \inf I'|>1}} 2\frac{\left|\mathbf{1}_I^T \Sigma_n^{-1} \mathbf{1}_{I'}\right|}{\sqrt{\tilde{\sigma}_n(I)\tilde{\sigma}_n(I')}} \exp\left[\frac{1}{2}\Delta_n^2 \sqrt{\tilde{\sigma}_n(I)\tilde{\sigma}_n(I')}\right] = o\left(\frac{1}{\lambda_n^2}\right),$$

which concludes the proof. □

Since Lemma 11 guarantees that Lemma 9 can be applied in the setting of Theorem 1, the proof of the latter now follows immediately from Lemmas 8 and 9.

Proof of Theorem 1. The two Lemmas 8 and 9 yield that the asymptotic detection boundary is (in terms of Δ_n) given by

$$(\sqrt{2}-\varepsilon_n)\sqrt{\frac{-\log\lambda_n}{n\lambda_n}}\sup_{I\in\mathcal{I}(\lambda_n)}\sqrt{\frac{n\lambda_n}{\tilde{\sigma}_n(I)}} \precsim \Delta_n \precsim (\sqrt{2}+\tilde{\varepsilon}_n)\sqrt{\frac{-\log\lambda_n}{n\lambda_n}}\inf_{I\in\mathcal{I}(\lambda_n)}\sqrt{\frac{\sigma_n(I)}{n\lambda_n}}, \tag{5.11}$$

as $n\to\infty$. For any $I\in\mathcal{I}(\lambda_n)$, it follows from Theorem 18.2.1 of Ibragimov and Linnik [36] (see Section 6.1 of Supplement A.2) that $\sigma_n(I)=n\lambda_n f(0)(1+o(1))$, and Lemma 10 yields

$$\tilde{\sigma}_n(I)=\frac{n\lambda_n}{f(0)}(1+o(1)),\quad n\to\infty.$$

Plugging this into (5.11) finishes the proof. □

Acknowledgments

Axel Munk and Frank Werner gratefully acknowledge financial support by the German Research Foundation DFG through subproject A07 of CRC 755, and Markus Pohlmann acknowledges support through RTG 2088. We are furthermore grateful to helpful comments of two anonymous referees and the associate editor.

Appendix A.2

Supplementary material to "Bump detection in the presence of dependency: Does it ease or does it load?"

Farida Enikeeva[1,2], Axel Munk[3,4,5], Markus Pohlmann[3] and Frank Werner[6]

[1]Laboratoire de Mathématiques et Applications, CNRS UMR 7348, Université de Poitiers, Site du Futuroscope - Téléport 2, 11 Boulevard Marie et Pierre Curie, Bâtiment H3 - TSA 61125, 86073 Poitiers Cedex 9, France
[2]The A. A. Kharkevich Institute for Information Transmission Problems, Russian Academy of Science, Bolshoy Karentny per. 19-1, 127051 Moscow, Russia
[3]Institute for Mathematical Stochastics, University of Göttingen, Goldschmidtstr. 7, 37077 Göttingen, Germany
[4]Felix-Bernstein-Institute for Mathematical Statistics in the Biosciences, University of Göttingen, Goldschmidtstr. 7, 37077 Göttingen, Germany
[5]Statistical Inverse Problems in Biophysics Group, Max Planck Institute for Biophysical Chemistry, Am Faßberg 11, 37077 Göttingen, Germany
[6]Institute for Mathematics, University of Würzburg, Emil-Fischer-Str. 30, 97074 Würzburg, Germany

Abstract

In this material we provide all remaining proofs and recall several results that are necessary for our theory.

6 Supplementary material and additional proofs

6.1 Auxiliary results

We begin this section by stating several useful results from various sources that we are going to use in our proofs.

Weak law of large numbers for arrays of dependent variables

Definition 12 (Sung, Lisawadi and Volodin [67]). *Let $\{X_{nk} : n \in \mathbb{N}, u_n \le k \le v_n\}$ be an array of random variables with $v_n - u_n \to \infty$ as $n \to \infty$. Additionally, let $r > 0$, and $(k_n)_{n\in\mathbb{N}}$ be a sequence of positive integers, such that $k_n \to \infty$ as $n \to \infty$.*
Let $(h(n))_{n\in\mathbb{N}}$ be a sequence of positive constants, such that $h(n) \nearrow \infty$ as $n \to \infty$. The array $\{X_{nk} : n \in \mathbb{N}, u_n \le k \le v_n\}$ is said to be h-integrable with exponent r if

$$\sup_{n\in\mathbb{N}} \frac{1}{k_n} \sum_{k=u_n}^{v_n} \mathbb{E}\left[|X_{nk}|^r\right] < \infty, \quad \text{and} \quad \lim_{n\to\infty} \frac{1}{k_n} \sum_{k=u_n}^{v_n} \mathbb{E}\left[|X_{nk}|^r \mathbb{1}\left\{|X_{nk}|^r > h(n)\right\}\right] = 0.$$

With this, we have the following.

Theorem 13 (Wang and Hu [71]). *Let m be a positive integer. Suppose that $\{X_{nk}, u_n \le k \le v_n,\ n \ge 1\}$ is an array of non-negative random variables with $\mathrm{Cov}(X_{nk}, X_{nk}) \le 0$ whenever $|j-k| \ge m$, $u_n \le j$, $k \le v_n$, for each $n \ge 1$ and is h-integrable with exponent $r = 1$ for a sequence $k_n \to \infty$ and $h(n) \uparrow \infty$, such that $h(n)/k_n \to 0$ as $n \to \infty$. Then*

$$\frac{1}{k_n} \sum_{k=u_n}^{v_n} (X_{nk} - EX_{nk}) \to 0$$

in L_1 and hence in probability, as $n \to \infty$.

Remark 14. *In fact, the original theorem in Wang and Hu [71] is slightly stronger, but Theorem 13 as stated above is sufficient for our purposes.*

Remark 15. *We can relax the condition $\mathrm{Cov}(X_{nj}, X_{nk}) \le 0$ whenever $|j-k| \ge m$, $u_n \le j, k \le v_n$ in Theorem 13 to requiring only that*

$$\limsup_{n\to\infty} \frac{1}{k_n^2} \sum_{\substack{j,k=u_n \\ |j-k|\ge m}}^{v_n} \mathrm{Cov}(X_{nj}, X_{nk}) \le 0.$$

Decay of precision matrices

The following result is due to Jaffard [41] and was used in [31] as a key tool in the analysis of a higher criticism test for detection of sparse signals observed in correlated noise. Here it is stated as it was formulated and proven in [31].

Lemma 16 (Hall and Jin [31]). *Let Σ_n, $n \ge 1$ be a sequence of $n \times n$ correlation matrices, such that $\|\Sigma_n\| \ge c > 0$, where $\|\Sigma_n\|$ is the operator norm of Σ_n as an operator from $\mathbb{R}^n$ to $\mathbb{R}^n$. If for some constants $\kappa > 0, C > 0$,*

$$|\Sigma_n(i,j)| \le C(1 + |i-j|)^{-(1+\kappa)},$$

then there is a constant $C' > 0$ depending on κ, C, and c, such that

$$|\Sigma_n^{-1}(i,j)| \le C'(1 + |i-j|)^{-(1+\kappa)}.$$

Long-run variance of partial sums of a stationary time series

Here we recall the well-known result given in Theorem 28.2.1 of Ibragimov and Linnik [36] on the explicit formula for the variance of the sum of n consecutive realizations of a stationary process. We adapt the notation to our case.

Suppose that $(X_n)_{n\in\mathbb{Z}}$ is a centered stationary sequence with the autocovariance function $\gamma(h)$, $h \in \mathbb{Z}$ and the spectral density $f(\nu)$, $\nu \in [-1/2, 1/2)$. Let $S_n = \sum_{i=1}^{n} X_i$.

Theorem 17 (Ibragimov and Linnik [36])**.** *The variance of S_n in terms of $\gamma(h)$ and $f(\nu)$ is given by*

$$\mathrm{Var}[S_n] = \sum_{|h|<n} (n-|h|)\gamma(h) = \int_{-1/2}^{1/2} \frac{\sin^2(\pi n\nu)}{\sin^2(\pi\nu)} f(\nu)\, d\nu.$$

If the spectral density $f(\nu)$ is continuous at $\nu = 0$, then

$$\mathrm{Var}[S_n] = f(0)n + o(n), \quad n \to \infty.$$

6.2 Remaining proofs for section 2

Proof of Lemma 10

Proof. We are inspired by the proof of Proposition C.1 in Keshavarz, Scott and Nguyen [45], that was dropped from the final paper [44], although we are able to make some simplifications, since a slightly weaker result suffices for our purposes. In addition, we use this opportunity to fix several minor inaccuracies in their proof.

Recall that $\mathcal{T}(f)$ is the infinite Toeplitz matrix generated by the spectral density f and that $\Sigma_n = \mathcal{T}_n(f)$ is the corresponding truncated Toeplitz matrix.

Let $\mathcal{T}(g)$ be the infinite Toeplitz matrix generated by $g = 1/f$, i.e. the matrix with elements $\mathcal{T}(g)(i,j) = g_{|i-j|}$, where $g_0, g_1, \ldots$ are the Fourier coefficients of g. Let $\mathcal{H}(f)$ and $\mathcal{H}(g)$ be the Hankel matrices generated by f and g, respectively, i.e. the matrices

$$\mathcal{H}(f) = \begin{pmatrix} f_1 & f_2 & f_3 & \ldots \\ f_2 & f_3 & f_4 & \ldots \\ f_3 & f_4 & f_5 & \ldots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \quad \text{and} \quad \mathcal{H}(g) = \begin{pmatrix} g_1 & g_2 & g_3 & \ldots \\ g_2 & g_3 & g_4 & \ldots \\ g_3 & g_4 & g_5 & \ldots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

It follows from Proposition 1.12 of Böttcher and Silbermann [8], that

$$\mathcal{T}(f)^{-1} = \mathcal{T}(g) + \mathcal{T}(f)^{-1}\mathcal{H}(f)\mathcal{H}(g).$$

Let v_I be the extension of the vector $\mathbf{1}_I$ to an element of l^2 by zero-padding. As in Keshavarz, Scott and Nguyen [44], from the above identity and the definition of the operator norm, we find

$$\begin{aligned} \left|v_I^T\mathcal{T}(f)^{-1}v_I - v_I^T\mathcal{T}(g)v_I\right| &= \left|\left\langle \mathcal{H}(f)\mathcal{T}(f)^{-1}v_I, \mathcal{H}(g)v_I\right\rangle\right| \\ &\le \|\mathcal{H}(f)\mathcal{T}(f)^{-1}\|\|v_I\|_{\ell_2}\|\mathcal{H}(g)v_I\|_{\ell_2} \\ &\le \|\mathcal{H}(f)\mathcal{T}(f)^{-1}\|\sqrt{n\lambda_n}\left[\sum_{\{r:v_I(r)=1\}} \|\mathcal{H}(g)e_r\|_{\ell_2}\right] \end{aligned}$$

where $e_r = (0,\ldots,0,1,0,\ldots)^T$ is the sequence whose r-th entry is 1, and $\|\mathcal{H}(f)\mathcal{T}(f)^{-1}\|$ is the operator norm of $\mathcal{H}(f)\mathcal{T}(f)^{-1}$ as an operator from ℓ^2 to ℓ^2. Since $\|\mathcal{T}(f)\| = \sup_{\nu\in[0,1)} f(\nu) < \infty$, we have $\|\mathcal{T}(f)^{-1}\| < \infty$ by the inverse mapping theorem. It follows that $\|\mathcal{H}(f)\mathcal{T}(f)^{-1}\| < \infty$, because clearly $\|\mathcal{H}(f)\| < \infty$. Since f is bounded away from 0, it is well known that the Fourier coefficients g_k, $k \in \mathbb{Z}$ of g decay at the same rate as the Fourier coefficients of f, i.e.

$$|g_k| \le C'(1+|k|)^{-(1+\kappa)},$$

for $k \in \mathbb{Z}$. Following Keshavarz, Scott and Nguyen [44]) we see that

$$\begin{aligned} \sum_{\{r:v_I(r)=1\}} \|\mathcal{H}(g)e_r\|_{\ell_2} = \sum_{\{r:v_I(r)=1\}} \left(\sum_{j=r}^{\infty} |g_j|^2\right)^{\frac{1}{2}} &\le \sum_{\{r:v_I(r)=1\}} \left(\int_r^{\infty} x^{-2(1+\kappa)}dx\right)^{\frac{1}{2}} \\ &\le C'' \sum_{\{r:v_I(r)=1\}} r^{-\left(\frac{1}{2}+\kappa\right)} \le C'' \sum_{r=1}^{\lfloor n\lambda_n\rfloor} r^{-\left(\frac{1}{2}+\kappa\right)}. \end{aligned}$$

It is then easy to see that the last expression is $O\left((n\lambda_n)^{\frac{1}{2}-\kappa}\right)$ if $\kappa < \frac{1}{2}$, and $O\left(\log(n\lambda_n)\right)$ if $\kappa = \frac{1}{2}$. Lastly, it is also clearly bounded if $\kappa > \frac{1}{2}$. Hence, in any of these cases it holds that

$$\sum_{\{r:v_I(r)=1\}} \|\mathcal{H}(g)e_r\|_{\ell_2} = o\left(\sqrt{n\lambda_n}\right).$$

Thus,

$$\left|v_I^T \mathcal{T}(f)^{-1} v_I - v_I^T \mathcal{T}(g) v_I\right| = o\left(n\lambda_n\right).$$

We now need to bound $v_I^T \mathcal{T}(g) v_I$. Let $(X_t)_{t\in\mathbb{N}}$ be a stationary random process with spectral density g. Then

$$v_I^T \mathcal{T}(g) v_I = \mathrm{Var}\left(\sum_{\{t:v_I(t)=1\}} X_t\right) = n\lambda_n(g(0) + o(1)),$$

as $n \to \infty$, where the last equality is due to Theorem 18.2.1 of Ibragimov and Linnik [36], see Section 5.3 of the Appendix for the precise statement of the theorem. (Note that g is continuous at 0 and $g(0) > 0$.) Thus,

$$v_I^T \mathcal{T}(f)^{-1} v_I = n\lambda_n(g(0) + o(1)).$$

Finally, by Theorem 2.11 of Böttcher and Grudsky [7], we have

$$\tilde{\sigma}_n(I) = v_I^T \Sigma_n^{-1} v_I = v_I^T \mathcal{T}(f)^{-1} v_I + \tilde{v}_I^T \left[\mathcal{T}(f)^{-1} - \mathcal{T}(g)\right] \tilde{v}_I + v_I^T D_n v_I,$$

where $\|D_n\| \to 0$, as $n \to \infty$, and $\tilde{v}_I$ arises from v_I through the transformation

$$\tilde{v}_I = (v_I(n), \ldots, v_I(1), 0, 0, \ldots).$$

As above, we have

$$\left|\tilde{v}_I^T \left[\mathcal{T}(f)^{-1} - \mathcal{T}(g)\right] \tilde{v}_I\right| = o\left(n\lambda_n\right),$$

and clearly, by Cauchy-Schwarz,

$$\left|v_I^T D_n v_I\right| \leq \|v_I\|^2 \|D_n\| = o\left(n\lambda_n\right).$$

This concludes the proof.

□

Remaining proofs

Proof of Theorem 2. Note that for any $1 \leq k \leq \lfloor \lambda_n^{-1} \rfloor$, under H_0, the random variables $\frac{\mathbf{1}_{I_k}^T \Sigma_n^{-1} Y}{\sqrt{\tilde{\sigma}_k}}$ with $\tilde{\sigma}_k$ as in (2.5) are identically distributed (dependent) standard Gaussian. Note that $\tilde{\sigma}_k = \tilde{\sigma}_n(I_k)$ with our former notation. The union bound and the elementary tail inequality $\mathbb{P}[|Z| > x] \leq 2e^{-x^2/2}$ for $Z \sim \mathcal{N}(0,1)$, yield

$$\begin{aligned}
\tilde{\alpha}(\Phi_n^{\mathrm{d}}) &= \mathbb{P}_0\left[T_n(Y) > c_{\alpha,n}\right] \\
&\leq \lfloor \lambda_n^{-1} \rfloor \sup_{1\leq k\leq \lfloor \lambda_n^{-1} \rfloor} \mathbb{P}_0\left[\frac{|\mathbf{1}_{I_k}^T \Sigma_n^{-1} Y|}{\sqrt{\tilde{\sigma}_k}} > c_{\alpha,n}\right] \\
&= \lfloor \lambda_n^{-1} \rfloor \mathbb{P}\left[|Z| > c_{\alpha,n}\right] \leq 2\lfloor \lambda_n^{-1} \rfloor \exp\left(-\frac{c_{\alpha,n}^2}{2}\right) \leq \alpha.
\end{aligned}$$

This proves $\tilde{\alpha}\left(\Phi_n^{\mathrm{d}}\right) \leq \alpha$ for all $n \in \mathbb{N}$.

Concerning the type II error, note that, under H_1, i.e. if $Y \sim \mathcal{N}(\delta_n \mathbf{1}_{I_k}, \Sigma_n)$ for some $k \in \{1, \ldots, \lfloor \lambda_n^{-1} \rfloor\}$, we have for all local test statistics on the right-hand side of (2.4) that

$$\frac{\mathbf{1}_{I_m}^T \Sigma_n^{-1} Y}{\sqrt{\sigma_m}} \sim \mathcal{N}\left(\frac{\delta_n \mathbf{1}_{I_m} \Sigma_n^{-1} \mathbf{1}_{I_k}}{\sqrt{\sigma_m}}, 1\right), \quad m = 1, \ldots, \lfloor \lambda_n^{-1} \rfloor.$$

Plugging in (2.6) and (2.8), it follows that the type II error satisfies

$$\begin{aligned}
\tilde{\beta}(\Phi_n^{\mathrm{d}}, \Sigma_n, \Delta_n, \lambda_n) &= \sup_{1\le k\le \lfloor \lambda_n^{-1}\rfloor} \sup_{|\delta_n|\ge \Delta_n} \mathbb{P}_{\delta_n,k}\left[\Phi_n^{\mathrm{d}}(Y)=0\right] \\
&= \sup_{1\le k\le \lfloor \lambda_n^{-1}\rfloor} \sup_{|\delta_n|\ge \Delta_n} \mathbb{P}\left[\sup_{1\le m\le \lfloor \lambda_n^{-1}\rfloor} \left| Z_m + \frac{\delta_n \mathbf{1}_{I_k}\Sigma_n^{-1}\mathbf{1}_{I_m}}{\sqrt{\sigma_m}}\right| \le c_{\alpha,n}\right] \\
&\le \sup_{1\le k\le \lfloor \lambda_n^{-1}\rfloor} \sup_{|\delta_n|\ge \Delta_n} \mathbb{P}\left[\left| Z_k + \delta_n\sqrt{\tilde{\sigma}_k}\right| \le c_{\alpha,n}\right] \\
&\le \sup_{1\le k\le \lfloor \lambda_n^{-1}\rfloor} \sup_{|\delta_n|\ge \Delta_n} \mathbb{P}\left[|\delta_n|\sqrt{\tilde{\sigma}_k} - |Z_k| \le c_{\alpha,n}\right] \\
&\le \mathbb{P}\left[|Z| > \Delta_n \inf_{1\le k\le \lfloor \lambda_n^{-1}\rfloor} \sqrt{\tilde{\sigma}_k} - c_{\alpha,n}\right]
\end{aligned}$$

which proves the claim. □

Proof of Corollary 3 . The claim follows directly from Theorem 2 and the standard Gaussian tail bound $\mathbb{P}\left[|Z|>x\right] \le 2e^{-x^2/2}$ via

$$\begin{aligned}
&\mathbb{P}\left[|Z| > \Delta_n \inf_{1\le k\le \lfloor \lambda_n^{-1}\rfloor} \sqrt{\tilde{\sigma}_k} - c_{\alpha,n}\right] \\
=&\mathbb{P}\left[|Z| > (1+\varepsilon_n)\sqrt{-2\log\lambda_n} - \sqrt{-2\log(\lambda_n) + 2\log(2\alpha^{-1})}\right] \\
\le&\mathbb{P}\left[|Z| > \varepsilon_n\sqrt{-2\log\lambda_n} - \sqrt{2\log(2\alpha^{-1})}\right] \\
\le&\exp\left(-\frac{1}{2}\left(\varepsilon_n\sqrt{-2\log\lambda_n} - \sqrt{2\log(2\alpha^{-1})}\right)^2\right) \\
\le&\exp\left(-\frac{1}{2}\left(\sqrt{2\log(\alpha^{-1})}\right)^2\right) = \alpha.
\end{aligned}$$

□

6.3 Proofs for Section 3

Proof of Theorem 4. It is well-known (see, for example, [63], Sections 3.3–3.4), that the autocovariance function γ of an ARMA process is exponentially decaying, i.e.

$$|\Sigma_n(i,j)| = |\gamma(i-j)| \le Ce^{-\kappa|i-j|},$$

for some $C>0$, some $\kappa>0$ and all $1\le i,j\le n$. Thus, Assumption 2 is satisfied, and Theorem 4 follows immediately from Theorem 1. □

Properties of the precision matrix of an AR(p) process

Let Z_t be a stationary AR(p) process defined in (3.5) and Σ_n be the covariance matrix of n consecutive realizations of Z_t. Then the precision matrix $\Sigma_n^{-1} = (\Sigma_n^{-1}(i,j))$, $i,j=1,\dots,n$ is a $n\times n$ symmetric $2p+1$-diagonal matrix with the upper-triangle elements given by (see [64])

$$\Sigma_n^{-1}(i,j) = \begin{cases} \sum_{t=0}^{i-1} \phi_t\phi_{t+j-i}, & 1\le i\le j\le p \\ \sum_{t=0}^{p+i-j} \phi_t\phi_{t+j-i}, & 1\le i\le n-p,\ \max(i,p+1)\le j\le i+p \\ \sum_{t=0}^{n-j} \phi_t\phi_{t+j-i}, & n-p+1\le i\le j\ge n \\ 0, & i+p<j\le n,\ i\le n-p. \end{cases} \tag{6.1}$$

Note that Σ_n^{-1} is symmetric with respect to both the main diagonal and the antidiagonal, so that $\Sigma_n^{-1}(i,j) = \Sigma_n^{-1}(j,i)$ and $\Sigma_n^{-1}(i,j) = \Sigma_n^{-1}(n+1-j, n+1-i)$.

We can see from (6.1) that Σ_n^{-1} has two symmetric blocks $L = (l_{ij})$ and $R = (r_{ij})$ of size p with the elements related as $l_{ij} = r_{p+1-i,p+1-j} = \Sigma_n^{-1}(i,j)$, $i,j = 1,\dots,p$ (red blocks in Fig. 6). The other elements of Σ_n^{-1} are constant on the diagonals and are given by $\Sigma_n^{-1}(i, i+k) = D_k$, $i = p-k+1$, $k = 1,\dots,p$ (blue parts of the matrix in Fig. 6), where

$$D_k = \sum_{t=0}^{p-k} \phi_t \phi_{t+k}, \quad k = 0,\dots,p.$$

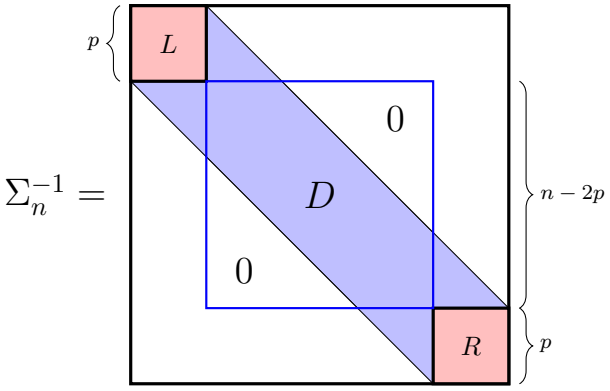


Figure 6: The matrix Σ_n^{-1} is symmetric $2p+1$-diagonal, the blocks L and R of size p are of size p, the blue part is has the same values D_k on the diagonals. The white part consists of zeros.

We are interested in the diagonal block sums of Σ_n^{-1} over the blocks of size r. We suppose that $1 \le r < \lfloor n/2 \rfloor - p$. The block sums of interest are

$$S_{r,m} = \mathbf{1}_{r,m}^T \Sigma_n^{-1} \mathbf{1}_{r,m}, \quad m = 1,\dots,n-r+1 \tag{6.2}$$

where $\mathbf{1}_{r,m} \in \mathbb{R}^n$ is the vector with entries

$$\mathbf{1}_{r,m}(i) = \begin{cases} 1, & i = m,\dots,m+r-1, \\ 0, & \text{otherwise} \end{cases}$$

Note that the key quantities $\tilde{\sigma}_k$ that appear in the lower and upper bounds of testing (5.3) and (2.8) are related to (6.2) as follows,

$$\tilde{\sigma}_k = S_{\lfloor n\lambda_n \rfloor, (k-1)\lfloor n\lambda_n \rfloor + 1}, \quad k = 1,\dots,\lfloor \lambda_n^{-1} \rfloor.$$

Lemma 18. *Suppose that $1 \le r \le n-2p$ and that $n \ge 3p$. The quantities $S_{r,m}$, $m = 1,\dots,n-r+1$ can be calculated directly using the following recursive formulas.*

1. *The first block sum is given by*

$$S_{r,1} = \begin{cases} \sum_{j=1}^{r} \left(\sum_{t=0}^{j-1} \phi_t \right)^2, & 1 \le r \le p, \\ \sum_{j=1}^{p} \left(\sum_{t=0}^{j-1} \phi_t \right)^2 + (r-p) \left(\sum_{t=0}^{p} \phi_t \right)^2, & p \le r \le n-p \end{cases} \tag{6.3}$$

2. If $r \le p$, then

$$S_{r,m+1} = S_{r,m} + \begin{cases} \left(\sum_{t=0}^{r-1} \phi_{t+i}\right)^2, & 1 \le m \le p+1-r \\ \left(\sum_{t=0}^{p-i} \phi_{t+i}\right)^2, & p+1-r \le m \le p \\ 0, & p+1 \le m \le n-p-r \\ -\left(\sum_{t=n-i-p}^{r-1} \phi_{n-i-t}\right)^2, & n-p-r+1 \le m \le n-p \\ -\left(\sum_{t=0}^{r-1} \phi_{n-i-t}\right)^2, & n-p \le m \le n-r \end{cases} \tag{6.4}$$

3. If $p \le r \le n-2p$, then

$$S_{r,m+1} = S_{r,m} + \begin{cases} \left(\sum_{t=0}^{p-i} \phi_{t+i}\right)^2, & 1 \le m \le p \\ 0, & p+1 \le im \le n-p-r \\ -\left(\sum_{t=n-i-p}^{r-1} \phi_{n-i-t}\right)^2, & n-p-r+1 \le m \le n-r. \end{cases} \tag{6.5}$$

The proof of the lemma is omitted. It follows from simple algebra and the relation

$$D_0 + 2\sum_{k=1}^{p} D_k = \sum_{t=0}^{p} \phi_t^2 + 2\sum_{k=1}^{p}\sum_{t=0}^{p-k} \phi_t \phi_{t+k} = \left(\sum_{t=0}^{p} \phi_t\right)^2.$$

Using the result of Lemma 18, we can calculate the constants $\tilde{\sigma}_k$.

Proof of Lemma 5. According to definition 6.2, the quantities $\tilde{\sigma}_k$ can be written as

$$\tilde{\sigma}_k = S_{\lfloor n\lambda_n \rfloor,(k-1)\lfloor n\lambda_n \rfloor+1}, \quad k = 1, \dots, \lfloor \lambda_n^{-1} \rfloor.$$

Note that it follows immediately from Lemma 18 that for any fixed $1 \le r \le n-2p$ the function $S_{r,m}$, $m = 1, \dots, n-r+1$ is monotone increasing for $m \le p+1$, constant for $p+1 \le m \le n-p-r+1$ and decreasing for $m \ge n-p-r+1$. Moreover, this function is symmetric in a sense that $S_{r,m} = S_{r,n-r-m+2}$, $m = 1, \dots, n-r+1$. Therefore, it follows that

$$\inf_{1 \le k \le \lfloor \lambda_n^{-1} \rfloor} \tilde{\sigma}_k = \min_{1 \le k \le \lfloor \lambda_n^{-1} \rfloor} S_{\lfloor n\lambda_n \rfloor,(k-1)\lfloor n\lambda_n \rfloor+1} = S_{\lfloor n\lambda_n \rfloor,1}$$

and

$$\sup_{1 \le k \le \lfloor \lambda_n^{-1} \rfloor} \sigma_k = \max_{1 \le k \le \lfloor \lambda_n^{-1} \rfloor} S_{\lfloor n\lambda_n \rfloor,(k-1)\lfloor n\lambda_n \rfloor+1} = S_{\lfloor n\lambda_n \rfloor,p+1}.$$

Note that the condition $\lfloor n\lambda_n \rfloor < n-2p$ will guarantee that the maximum is attained at the interval where the function S is constant (for some k that satisfies $p+1 \le (k-1)\lfloor n\lambda_n \rfloor + 1 \le n-p-r+1$) and, consequently, will be equal to $S_{\lfloor n\lambda_n \rfloor,p+1}$.

We obtain the statement of the lemma applying the recursive formulas of Lemma 18. □

Appendix B

Minimax detection of localized signals in statistical inverse problems

Markus Pohlmann[1], Frank Werner[2] and Axel Munk[1,3,4,5]

[1]Institute for Mathematical Stochastics, University of Göttingen, Goldschmidtstr. 7, 37077 Göttingen, Germany
[2]Institute for Mathematics, University of Würzburg, Emil-Fischer-Str. 30, 97074 Würzburg, Germany
[3]Felix-Bernstein-Institute for Mathematical Statistics in the Biosciences, University of Göttingen, Goldschmidtstr. 7, 37077 Göttingen, Germany
[4]Max Planck Institute for Biophysical Chemistry, Am Faßberg 11, 37077 Göttingen, Germany
[5]Cluster of Excellence MBExC: "Multiscale Bioimaging: From Molecular Machines to Networks of Excitable Cells", University Medical Center, Robert-Koch-Str. 40, 37075 Göttingen, Germany

Abstract

We investigate minimax testing for detecting local signals or linear combinations of such signals when only indirect data is available. Naturally, in the presence of noise, signals that are too small cannot be reliably detected. In a Gaussian white noise model, we discuss upper and lower bounds for the minimal size of the signal such that testing with small error probabilities is possible. In certain situations we are able to characterize the asymptotic minimax detection boundary. Our results are applied to inverse problems such as numerical differentiation, deconvolution and the inversion of the Radon transform.

Keywords: Hypothesis testing, minimax signal detection, statistical inverse problems.
AMS classification numbers: 62F03, 65J22, 65T60, 60G15.

1 Introduction

In many practical applications one aims to infer on properties of a quantity which is not directly observable. As a guiding example, consider computerized tomography (CT), where the interior (more precisely the tissue density) of the human body is imaged via the absorption of X-rays along straight lines. Mathematically, the relation between the available measurements Y (absorption

along lines, the so-called sinogram) and the unknown quantity of interest f (the tissue density) is described by the Radon transform, which is an integral operator to be described in more detail later (cf. Figure 1 for illustration). Potential further applications include astronomical image processing, magnetic resonance imaging, non-destructive testing and super-resolution microscopy, to mention a few. Typically, the measurements are either of random nature themselves (as e.g. in positron emission tomography (PET, see [70]), magnetic resonance imaging (MRI, see [46]) or super-resolution microscopy (see [54])) and/or additionally corrupted by measurement noise. This motivates us to consider the inverse Gaussian white noise model

$$Y_\sigma = Af + \sigma\xi \tag{1.1}$$

with a (known) bounded linear operator $A : \mathcal{X} \to \mathcal{Y}$ mapping between (real or complex) Hilbert spaces $\mathcal{X}$ and $\mathcal{Y}$, noise level $\sigma > 0$ and a Gaussian white noise ξ on $\mathcal{Y}$ (details will be given in section 2).

A major effort of research is devoted to the development and analysis of estimation and recovery methods of the signal f from the measurements Y_σ (see Section 1.2 for some references). However, when f is expected to be very close to some reference f_0, by which we mean that either $f = f_0$ or f deviates from f_0 by only a few localized components (anomalies), then instead of full recovery of f, one might be more interested in testing whether $f = f_0$ or not. This is especially relevant, since, when the signal-to-noise level is too small for full recovery, then testing may still be informative as it is well-known to be a simpler task (see e.g. [61] and the references therein). Although of practical importance, testing in model (1.1) is a much less investigated endeavor than estimation and a full theoretical understanding has not been achieved yet. Hence, in this paper, we are interested in analyzing such testing methodology for inferring on f based on the available data Y_σ. Note that, due to the linearity of the model (1.1), we can w.l.o.g. assume that $f_0 = 0$. Thus, we suppose that either $f = 0$ (no anomaly is present) or $f = \delta u$ (an anomaly given by δu is present), where $u \in \mathcal{F}_\sigma$ for some (finite) class $\mathcal{F}_\sigma \subseteq \mathcal{X}$ of non-zero functions, that are – in some sense – normalized, and the constant factor δ describes its orientation, and – more importantly – how "large" or "pronounced" the signal f is. We consider the family of testing problems

$$H_0 : f = 0 \quad \text{against} \quad H_{1,\sigma} : f = \delta u \text{ for some } u \in \mathcal{F}_\sigma \text{ and } |\delta| > \mu_\sigma, \tag{1.2}$$

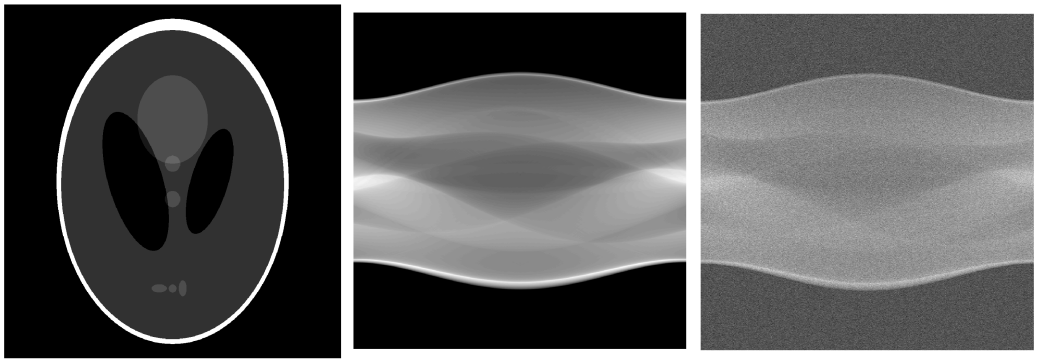
where $(\mu_\sigma)_{\sigma>0}$ is a family of non-negative real numbers. This can be viewed as the problem of *detecting* an anomaly from the set $\{\delta u : u \in \mathcal{F}_\sigma, |\delta| > \mu_\sigma\}$.

We suppose that the family of classes $(\mathcal{F}_\sigma)_{\sigma>0}$ is chosen in advance. This choice is crucial for the analysis of the problem and it depends solely on the specific application: For CT we might think of small inclusions such as tumors, cf. Figure 1, where certain wavelets are used as mathematical representation. If no a priori knowledge about potential anomalies is known, it is natural to start by considering dictionaries $(u_k)_{k\in I}$ with good expressibility in $\mathcal{X}$, e.g. frames or wavelets, and set $\mathcal{F}_\sigma = \{u_k : k \in I_\sigma\}$ for subsets I_σ of I. The particular choices that we analyze in this paper will be built from such dictionaries, see also [21] and [34] for recent references in the context of estimation.

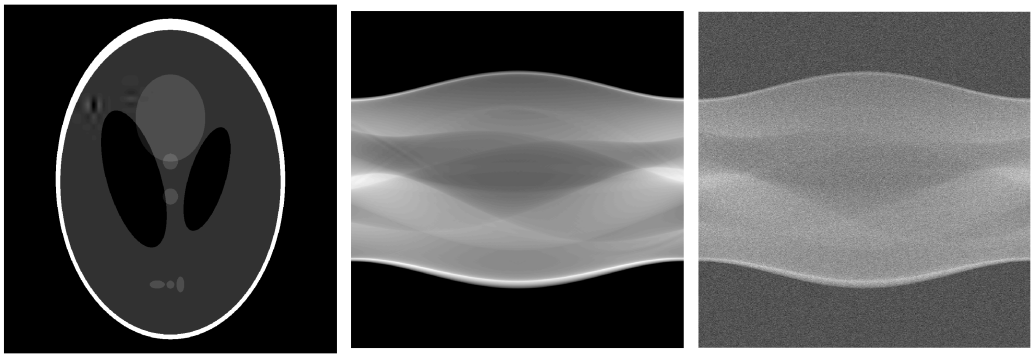
1.1 Aim of the paper

Given a family of classes $(\mathcal{F}_\sigma)_{\sigma>0}$, our main objective will be to assess to what extent powerful tests for the testing problem (1.2) exist. The answer will usually depend on the size of μ_σ: If μ_σ is large enough, then powerful tests exist, and if μ_σ is too small, then no test has high power. Hence, we aim to find a minimal family of thresholds $(\mu_\sigma^*)_{\sigma>0}$, such that powerful detection at a controlled error rate is still possible. Vice versa, such a minimal family would determine which signals can not be detected reliably, even when they are present.

To this end, we extend the existing theory on minimax signals detection in inverse problems focusing on localized signals and linear combinations of localized signals, which are common in practice. This has, to the best of our knowledge, not been investigated yet. We present upper bounds, lower bounds and asymptotics for the minimal values of μ_σ such that powerful tests for



(a) Reference image (b) Sinogram of (a) (c) Noisy sinogram of (a)



(d) Distorted Image (e) Sinogram of (d) (f) Noisy sinogram of (d)

Figure 1: Illustration of structured hypothesis testing in the CT example. To infer whether the unknown signal deviates from a reference image, we use a test based on the noisy sinogram. In the above example, when the distortion is assumed to be a linear combination of certain wavelets (cf. Sections 3.2 and 4), then the results of Theorem 3.9 imply the existence of a test which is able to distinguish the distorted (1d) from the undistorted image (1a) with type I and type II error both at most 0.05, based on the measurements 1f.

testing problems given by (1.2) exist. They depend on the difficulty of the inverse problem induced by the forward operator A, the cardinality of I_σ (denoted by $|I_\sigma|$) and the inner products between the images Au_k, $k \in I_\sigma$, of the potential anomalies. We stress that our results can be applied to a variety of dictionaries $(u_k)_{k\in I}$, such as wavelets, whereas previous results were restricted to dictionaries based on the SVD of the operator A. As one particular example, our results can be applied to the situation where the dictionary $(u_k)_{k\in I}$ is (a subset of) the famous Wavelet-Vaguelette-decomposition (WVD, see [18]) or the Vaguelette-Wavelet-decomposition (VWD, see [1]) of A.

Figure 1 serves as an illustrative example. If it is known a priori, that the anomaly which distorts the reference image is a linear combination of a certain collection of wavelets (see the discussion in Sections 3.2 and 4 for details), then our results suggest that the anomaly that is present in 1d is large enough, such that there is a test which is able to distinguish the distorted (1d) from the undistorted image (1a) with type I and type II error both at most 0.05, based on the measurements 1f (see Theorem 3.9). Note that our results are not restricted to wavelets. In fact, most of our results are applicable under very mild conditions on the dictionary $(u_k)_{k\in I}$.

We stress that this paper does not constitute an exhaustive study of the subject. Rather, we aim to provide some first analysis and discuss some illustrative examples.

1.2 Connection to existing literature

As the literature on estimating f in model (1.1) is vast (see e.g. [22], [32], [6], [18], [1], [2], [72], [21]), we confine ourselves to briefly reviewing the literature on (minimax) testing theory, the topic of the present paper.
First of all, there is extensive literature about minimax signal detection for the direct problem, i.e. when $\mathcal{X} = \mathcal{Y}$ and A is the identity. We only mention the seminal works [38] and [40]. Usually, the hypothesis "$f = 0$" is tested against alternatives of the form "$f \in \mathcal{F}$ and $\|f\|_{\mathcal{X}} > \mu_\sigma$", where $\mathcal{F}$ is a certain class of functions, for example defined by certain smoothness conditions. The indirect case where A is allowed to differ from the identity has e.g. been treated in [48], [39], [37], [53], [4]. Note, that our testing problem (1.2) has an alternative which is substantially different to testing against a smoothness condition $f \in \mathcal{F}$ with sufficiently large norm. Our approach is different, as instead of e.g. smoothness constraints, expressed through $\mathcal{F}$, we consider the alternative that f is an element of a very specific set of candidate functions. We refer to [25] and [23], where systems of scaled and translated rectangle functions (bumps) in a direct setting were considered.
Finally, we want to highlight [49] explicitly as they consider alternatives consisting of linear combinations of anomalies given in terms of the SVD of the operator A, which served as a point of reference and inspiration to parts of this study.

1.3 Outline

We start by giving a detailed description about our model and some basic facts about testing and minimax signal detection in section 2. Section 3 contains the main results: In section 3.1 we assume that $\mathcal{F}_\sigma$ is a collection of frame elements, and in section 3.2 we assume that $\mathcal{F}_\sigma$ contains functions in the linear span of a collection of frame elements. Both sections also include discussions about conditions that frames need to satisfy for our results to be applicable. We present illustrative simulation studies in section 4. All proofs are postponed to section 6.

2 Preliminaries

2.1 Detailed model assumptions

The model (1.1) has to be understood in a weak sense, i.e.

$$Y_\sigma(h) = \langle Af, h\rangle_{\mathcal{Y}} + \sigma\xi(h), \quad h \in \mathcal{Y}. \tag{2.1}$$

The error ξ is a Gaussian white noise on $\mathcal{X}$:

(1) If $\mathcal{X}$ and $\mathcal{Y}$ are real Hilbert spaces, we suppose that $\xi : \mathcal{Y} \to L^2(\Omega, P)$, for some some probability space $(\Omega, \mathcal{A}, P)$, is a linear mapping satisfying $\xi(h) \sim \mathcal{N}(0, \|h\|_{\mathcal{Y}}^2)$ and $\mathbb{E}\left(\xi(h)\xi(h')\right) = \langle h, h'\rangle_{\mathcal{Y}}$ for all $h, h' \in \mathcal{Y}$.

(2) If $\mathcal{X}$ and $\mathcal{Y}$ are complex Hilbert spaces, instead we suppose that $\xi(h) \sim \mathcal{CN}(0, 2\|h\|_{\mathcal{Y}}^2)$ and $\mathbb{E}(\xi(h)\overline{\xi(h')}) = 2\langle h, h'\rangle_{\mathcal{Y}}$. Here $X \sim \mathcal{CN}(0,1)$ means that X is distributed according to the standard complex normal distribution, i.e. $X = X_1 + iX_2$, where $X_1, X_2 \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1/2)$.

We will use the notation $\langle Y_\sigma, h\rangle_{\mathcal{Y}} := Y_\sigma(h)$ for convenience.

2.2 Notation

For a complex number z, we denote its real and imaginary part by $\Re z$ and $\Im z$, respectively.
For two families $(a_\sigma)_{\sigma>0}$, $(b_\sigma)_{\sigma>0}$ of non-negative real numbers we write $a_\sigma \precsim b_\sigma$ if $\limsup_{\sigma\to 0} a_\sigma/b_\sigma \leq 1$, and we write $a_\sigma \succsim b_\sigma$ if $\liminf_{\sigma\to 0} a_\sigma/b_\sigma \geq 1$. If $\lim_{\sigma\to 0} a_\sigma/b_\sigma = 1$, we write $a_\sigma \asymp b_\sigma$, and if $\lim_{\sigma\to 0} a_\sigma/b_\sigma = c < \infty$, we write $a_\sigma \sim b_\sigma$.

2.3 Testing and distinguishability

In the above testing problem (1.2), we wish to test the *hypothesis* H_0 against the *alternative* $H_{1,\sigma}$, which means making an educated guess (based on the data) about the correctness of the hypothesis when compared to the alternative, while keeping the error of wrongly deciding against H_0 under control. Tests are based on *test statistics*, i.e. measurable functions of the data Y_σ. We suppose that any test statistics can be expressed in terms of the Gaussian sequence $y_\sigma = (y_{\sigma,i})_{i\in\mathbb{N}}$ given by

$$y_{\sigma,i} := \langle Y_\sigma, e_i\rangle_{\mathcal{Y}} = \langle Af, e_i\rangle_{\mathcal{Y}} + \sigma\xi_i, \quad i \in \mathbb{N}, \tag{2.2}$$

where $\{e_i : i \in \mathbb{N}\}$ is a basis of the Hilbert space $\mathcal{Y}$, and, consequently, $\xi_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ (in the real case) or $\xi_i \overset{\text{i.i.d.}}{\sim} \mathcal{CN}(0,1)$ (in the complex case) for $i \in \mathbb{N}$. In the following, we use the notation Y_σ interchangeably for either the random process given by (2.1) or the random sequence given by (2.2), since they are equivalent in terms of the data they provide.
A test for the testing problem (1.2) can now be viewed as a measurable function of the sequence y_σ given by

$$\phi : \mathbb{K}^{\mathbb{N}} \to \{0,1\},$$

where $\mathbb{K}$ is either $\mathbb{R}$ or $\mathbb{C}$. The test ϕ can be understood as a decision rule in the following sense: If $\phi(y_\sigma) = 0$, the hypothesis is *accepted*. If $\phi(y_\sigma) = 1$, the hypothesis is *rejected* in favor of the alternative.
If H_0 is true, i.e. $f = 0$, but $\phi(y_\sigma) = 1$, we call this a *type I error* (the hypothesis is rejected although it is true). The probability to make a type I error is

$$\alpha_\sigma(\phi) := \mathbb{P}_0(\phi(y_\sigma) = 1),$$

where $\mathbb{P}_0$ denotes the distribution of y_σ given that H_0 is true. Likewise, the alternative might be true, but $\phi(y_\sigma) = 0$. We call this a *type II error* (the hypothesis is accepted although the alternative is true). Let us, for simplicity, introduce the notation $\mathcal{F}_\sigma(\mu_\sigma) = \{\delta u : u \in \mathcal{F}_\sigma, |\delta| \geq \mu_\sigma\}$. The type II error probability, given that a specific $f \in \mathcal{F}_\sigma(\mu_\sigma)$ is the true signal, is denoted as

$$\beta_\sigma(\phi, f) := \mathbb{P}_f(\phi(y_\sigma) = 0), \quad f \in \mathcal{F}_\sigma(\mu_\sigma),$$

where $\mathbb{P}_f$ denotes the distribution of y_σ given that f is the true underlying signal. Since the alternative is – in general – composite, i.e. does not only consist of only one element, the type II error probability will in general depend on the element f. For such composite alternatives we consider the worst case error given by the maximum type II error probability over $\mathcal{F}_\sigma(\mu_\sigma)$ for our analysis.
We say that the hypothesis H_0 is asymptotically *distinguishable* (in the minimax sense) from the family of alternatives $(H_{1,\sigma})_{\sigma>0}$ when there exist tests for the testing problems "H_0 against $H_{1,\sigma}$", $\sigma > 0$, that have both small type I and small maximum type II error probabilities. We define

$$\gamma_\sigma = \gamma_\sigma(\mu_\sigma) := \inf_{\phi\in\Phi_\sigma} \left[\alpha_\sigma(\phi) + \sup_{f\in\mathcal{F}_\sigma(\mu_\sigma)} \beta_\sigma(\phi, f)\right],$$

where Φ_σ is the set of all tests for the testing problem "H_0 against $H_{1,\sigma}$". In terms of γ_σ we say that H_0 and $H_{1,\sigma}$ are *distinguishable* if $\gamma_\sigma \to 0$, as $\sigma \to 0$. If $\gamma_\sigma \to 1$, we say that they are *indistinguishable*. We refer to [39] for an in-depth treatment.
For prescribed families $\mathcal{F}_\sigma$, we are interested in determining the smallest possible values μ_σ, such that H_0 and $H_{1,\sigma}$ are still asymptotically distinguishable, if possible. If a family $(\mu^*_\sigma)_{\sigma>0}$ exists, that satisfies

$$\gamma_\sigma(\mu_\sigma) \to 0 \quad \text{if} \quad \mu_\sigma \gtrsim \mu^*_\sigma, \quad \text{and} \quad \gamma_\sigma(\mu_\sigma) \to 1 \quad \text{if} \quad \mu_\sigma \precsim \mu^*_\sigma,$$

as $\sigma \to 0$, we call $(\mu^*_\sigma)_{\sigma>0}$the *(asymptotic) minimax detection boundary*. We may say that $(\mu^*_\sigma)_{\sigma>0}$ separates *detectable* and *undetectable* signals.

It is, however, not always possible to find such a sharp threshold. If the family $(\mu_\sigma^*)_{\sigma>0}$ only satisfies the weaker conditions

$$\gamma_\sigma(\mu_\sigma) \to 0 \quad \text{if} \quad \mu_\sigma/\mu_\sigma^* \to \infty, \quad \text{and} \quad \gamma_\sigma(\mu_\sigma) \to 1 \quad \text{if} \quad \mu_\sigma/\mu_\sigma^* \to 0,$$

we call it the *separation rate* of the family of testing problems "H_0 against $H_{1,\sigma}$".

Remark: Although we are mostly interested in the asymptotics of the problem, we will also state non-asymptotic results, which we deem interesting.

3 Results

Throughout the rest of the paper, we will assume that $(u_k)_{k\in I}$ is a countable collection of functions in $\mathcal{X}$, and $(I_\sigma)_{\sigma>0}$ is a family of finite subsets of I.

3.1 Alternatives given by finite collections of functions

We first suppose that $\mathcal{F}_\sigma$ consists of the appropriately normalized functions u_k, $k \in I_\sigma$, i.e. $\mathcal{F}_\sigma = \big\{\|Au_k\|_{\mathcal{Y}}^{-1} u_k : k \in I_\sigma\big\}$. As above, we write $\mathcal{F}_\sigma(\mu_\sigma) = \big\{\delta\|Au_k\|_{\mathcal{Y}}^{-1} u_k : k \in I_\sigma, |\delta| \geq \mu_\sigma\big\}$, so that testing problem (1.2) can be written as

$$H_0 : f = 0 \quad \text{against} \quad H_{1,\sigma} : f \in \mathcal{F}_\sigma(\mu_\sigma). \tag{3.1}$$

An upper bound for the detection boundary μ_σ^*

Any family of tests $(\phi_\sigma)_{\sigma>0}$ for the family of testing problems (3.1) yields an upper bound for μ_σ^*. It seems natural to choose maximum likelihood type tests as candidates, which are given by

$$\phi_{\sigma,\alpha}(y_\sigma) = \mathbb{1}\left\{\sup_{k\in I_\sigma} \frac{|\langle Y_\sigma, Au_k\rangle_{\mathcal{Y}}|}{\sigma\|Au_k\|_{\mathcal{Y}}} > c_{\alpha,\sigma}\right\}, \quad \sigma > 0, \tag{3.2}$$

for a given significance level $\alpha \in (0,1)$, and for appropriately chosen thresholds $c_{\alpha,\sigma}$ (which depend on whether the spaces $\mathcal{X}$ and $\mathcal{Y}$ are real or complex Hilbert spaces).

Theorem 3.1. *Let $N_\sigma = |I_\sigma|$ and assume that $N_\sigma \to \infty$, as $\sigma \to 0$. In addition, assume that*

$$\mu_\sigma \gtrsim (1+\varepsilon_\sigma)\sqrt{2\sigma^2 \log N_\sigma},$$

where $\varepsilon_\sigma \to 0$ and $\varepsilon_\sigma\sqrt{\log N_\sigma} \to \infty$ as $\sigma \to 0$. Then $\gamma_\sigma(\mu_\sigma) \to 0$ and thus, $\mu_\sigma^* \lesssim (1+\varepsilon_\sigma)\sqrt{2\sigma^2 \log N_\sigma}$.

The bound given in Theorem 3.1 does not depend on A and it depends on set of anomalies $(u_k)_{k\in I}$ and the family of candidate indices $(I_\sigma)_{\sigma>0}$ only through the cardinality N_σ. Thus, Theorem 3.1 has the advantage that it is (almost) always applicable, but it might be not very well suited for specific applications. We will see examples, where the bound is essentially sharp, and an example, where it is basically useless.

A lower bound for μ_σ^*

Theorem 3.2. *Let*

$$N_\sigma^* = \sup\{\#S : S \subseteq I_\sigma,\ \Re(\langle Au_k, Au_{k'}\rangle_{\mathcal{Y}}) \leq 0 \ \text{ for any two distinct } k, k' \in S\},$$

and assume that $N_\sigma^* \to \infty$. In addition, assume that

$$\mu_\sigma \lesssim (1-\varepsilon_\sigma)\sqrt{2\sigma^2 \log N_\sigma^*}, \tag{3.3}$$

where $(\varepsilon_\sigma)_{\sigma>0}$ is a family of positive real numbers such that $\varepsilon_\sigma \to 0$ and $\varepsilon_\sigma\sqrt{\log N_\sigma^*} \to \infty$ as $\sigma \to 0$. Then $\gamma_\sigma(\mu_\sigma) \to 1$ and thus, $\mu_\sigma^* \gtrsim (1-\varepsilon_\sigma)\sqrt{2\sigma^2 \log N_\sigma^*}$.

This theorem can be proven by using Proposition 4.10 and Lemma 7.2 of [40]. However, we will provide a self-contained and simple proof employing a weak law of large numbers for dependent random variables in section 6.
Theorem 3.2 implies that the number N_σ^* of negatively correlated image elements Au_k is the relevant quantity which determines the difficulty of testing (1.2). The actual cardinalty N_σ plays no role in the lower bound (3.3).

The detection boundary

As a consequence, we are now in position to describe the asymptotic detection boundary precisely in several situations. First, a combination of the previous theorems yields the following:

Corollary 3.3. *Assume that $N_\sigma = |I_\sigma| \to \infty$, and let*

$$M_\sigma = \sup_{k \in I_\sigma} \#\{k' \in I_\sigma : \Re(\langle Au_k, Au_{k'} \rangle_{\mathcal{Y}}) > 0\},$$

and assume that $M_\sigma N_\sigma^{-\varepsilon_\sigma} \to 0$ for a family $(\varepsilon_\sigma)_{\sigma>0}$ that satisfies $\varepsilon_\sigma \to 0$ and $\varepsilon_\sigma \sqrt{\log N_\sigma} \to \infty$ as $\sigma \to 0$. Then $\mu_\sigma^* \asymp \sqrt{2\sigma^2 \log N_\sigma}$.

In particular, Corollary 3.3 yields the asymptotic detection boundary, when $(Au_k)_{k \in I_\sigma}$ is orthogonal. Note that the assumptions of Corollary 3.3 are satisfied when M_σ is constant as $\sigma \to 0$. This has several applications, as we will see e.g. in Section 3.1.
Assume now that the operator $A : \mathcal{X} \to \mathcal{Y}$ is compact and has a singular value decomposition given by orthonormal systems $(\zeta_i)_{i \in \mathbb{N}}$ and $(\eta_i)_{i \in \mathbb{N}}$ in $\mathcal{X}$ and $\mathcal{Y}$, respectively, and singular values $(s_i)_{i \in \mathbb{N}}$.

Corollary 3.4. *Let $I = \mathbb{N}$ and $u_k = \zeta_k$ and $a_k = 1/s_k$ for $k \in \mathbb{N}$, and let $(I_\sigma)_{\sigma>0}$ be any family of finite subsets of I, such that $N_\sigma = |I_\sigma| \to \infty$, as $\sigma \to 0$. Then $\mu_\sigma^* \asymp \sqrt{2\sigma^2 \log N_\sigma}$.*

Remark: The detection thresholds for the SVD are clearly very easy to find, and could be deduced from other known results (see [39] for example). We include it here, since, as far as we know, it has not been stated explicitly before.

Frame decompositions

We have seen that sharp detection thresholds for the SVD can easily be found, but this does (usually) not cover the situation when we are interested in local anomalies. We will thus focus on other options for anomaly systems, particularly frames, for which be briefly introduce the most important notation. Let $\mathcal{H}$ be a separable Hilbert space, and let I be a countable index set. A sequence $(e_k)_{k \in I} \subseteq \mathcal{H}$ is called a frame of $\mathcal{H}$ if there exist constants $C_1, C_2 > 0$, such that for any $h \in \mathcal{H}$

$$C_1 \|h\|_{\mathcal{H}}^2 \leq \sum_{k \in I} |\langle h, e_k \rangle_{\mathcal{H}}|^2 \leq C_2 \|h\|_{\mathcal{H}}^2.$$

Since frames not have to be orthonormal, they provide great flexibility. Theorems 3.2 and 3.1 clearly apply to testing (1.2) with $u_k = e_k$, however, the fact that $(u_k)_{k \in I}$ constitutes a frame is, on its own, not enough to guarantee that we obtain a sharp detection boundary from Corollary 3.3.
In the following we show how frames $(u_k)_{k \in I}$ can be constructed, for which Corollary 3.3 can be applied. The idea is as follows: Since the bounds for the detection threshold mostly depend on properties of the images Au_k in $\mathcal{Y}$, we will simply start by defining a frame $(v_k)_{k \in I}$ in $\mathcal{Y}$ that will guarantee that the needed properties are satisfied, and then construct the corresponding frame $(u_k)_{k \in I}$ in $\mathcal{X}$, such that the pair $(u_k)_{k \in I}$, $(v_k)_{k \in I}$ is a decomposition of the operator A, and such that the assumptions of Corollary 3.3 are satisfied for any family of subsets $(I_\sigma)_{\sigma>0}$.

Assumption 3.5. *(i) There is a dense subspace $\tilde{\mathcal{Y}} \subseteq \mathcal{Y}$ with inner product $\langle \cdot, \cdot \rangle_{\tilde{\mathcal{Y}}}$ and norm $\| \cdot \|_{\tilde{\mathcal{Y}}}$, and constants $c_1, c_2 > 0$, such that*

$$c_1 \|x\|_{\mathcal{X}} \leq \|Ax\|_{\tilde{\mathcal{Y}}} \leq c_2 \|x\|_{\mathcal{X}}, \tag{3.4}$$

for all $x \in \mathcal{X}$.

(ii) There is a frame $(v_k)_{k \in I}$ of $\mathcal{Y}$ and a sequence $(\lambda_k)_{k \in I}$ of real numbers with $\alpha_k \neq 0$, and constants $a_1, a_2 > 0$, such that

$$a_1 \|y\|_{\tilde{\mathcal{Y}}}^2 \leq \sum_{k \in I} \lambda_k^2 |\langle y, v_k \rangle_{\mathcal{Y}}|^2 \leq a_2 \|y\|_{\tilde{\mathcal{Y}}}^2,$$

for all $y \in \tilde{\mathcal{Y}}$.

Assumption 3.5 implies that A as an operator from $\mathcal{X}$ to $\mathrm{ran}(A) \subseteq \tilde{\mathcal{Y}}$ is invertible. Now let $(v_k)_{k \in I}$ be a frame of $\mathrm{ran}(A)$ as in (ii). We apply the Gram-Schmidt procedure with respect to the inner product $\langle \cdot, \cdot \rangle_{\mathcal{Y}}$ to $(v_k)_{k \in I}$. This results in a sequence $(v_k^*)_{k \in I}$, which is a frame in $\mathrm{ran}(A)$ and which is orthogonal with respect to $\langle \cdot, \cdot \rangle_{\mathcal{Y}}$. Now we define

$$u_k = \lambda_k A^{-1} v_k^*,$$

for $k \in I$. The system $(u_k)_{k \in I}$ clearly yields sharp detection thresholds, as for any subset $I_\sigma \subset I$ it holds that $N_\sigma = N_\sigma^*$ by construction. Furthermore, it is a frame in $\mathcal{X}$, since for $x \in \mathcal{X}$

$$\sum_{k \in I} |\langle x, u_k \rangle_{\mathcal{X}}|^2 = \sum_{k \in I} \lambda_k^2 |\langle (A^*)^{-1} x, v_k^* \rangle_{\mathcal{Y}}|^2 \sim \|(A^*)^{-1} x\|_{\tilde{\mathcal{Y}}},$$

and

$$\|A^*\|_{\tilde{\mathcal{Y}} \to \mathcal{X}}^{-1} \|x\|_{\mathcal{X}} \leq \|(A^*)^{-1} x\|_{\tilde{\mathcal{Y}}} \leq \|(A^*)^{-1}\|_{\mathcal{X} \to \mathcal{X}} \|x\|_{\mathcal{X}}.$$

As a consequence we obtain the following.

Theorem 3.6. *Suppose that Assumption (3.5) is satisfied. Then for any frame $(u_k)_{k \in I}$ of $\mathcal{X}$, constructed as above, and for any family of subsets of indices $(I_\sigma)_{\sigma > 0}$ with $N_\sigma := |I_\sigma| \to \infty$ as $\sigma \to 0$, we have $\mu_\sigma^* \asymp \sqrt{2\sigma^2 \log N_\sigma}$.*

Examples

We discuss several commonly used operators and present a few typical examples of collections $(u_k)_{k \in I}$, for which the above theorems may or may not apply.

Integration

Let $\mathcal{X} = \mathcal{Y} = L^2(\mathbb{R})$ and let $A : \mathcal{X} \to \mathcal{Y}$ be the linear Fredholm integral operator given by

$$(Af)(x) = \int_{-\infty}^{x} f(t) dt, \quad x \in \mathbb{R},$$

for $f \in \mathcal{X}$. Suppose that ψ is a (mother) wavelet in $L^2(\mathbb{R})$, that satisfies $\int_{\mathbb{R}} \psi(x) dx = 0$, and for which the collection $(\psi_{j,k})_{j,k \in \mathbb{Z}}$ given by

$$\psi_{j,l}(x) = 2^{j/2} \psi(2^j x - l)$$

forms an orthogonal frame of $L^2(\mathbb{R})$. For an in-depth treatment of wavelet theory, we refer to [52] or [16].

Let us suppose that the system of possible anomalies is given by this wavelet system, i.e. we consider $\{u_{(j,l)} : (j,l) \in I = \mathbb{Z}^2\}$ with $u_{(j,l)} = \psi_{j,l}$. Assume further that ψ is compactly supported

with support size L, which implies that for any pair of indices (j,l) the number of indices k', such that $\operatorname{supp} u_{(j,l)} \cap \operatorname{supp} u_{(j,l')} \neq \emptyset$ is at most L.
Since, in practical applications, we would not expect to be able to obtain observations on the whole plane $\mathbb{R}^2$, we suppose that an anomaly, if one exists, must lie within some compact subset of $\mathbb{R}^2$, e.g. the unit interval $[0,1]$. For some family of integers $(j_\sigma)_{\sigma>0}$ that satisfies $j_\sigma \to \infty$ as $\sigma \to 0$ we define the family $(I_\sigma)_{\sigma>0}$ of "candidate" indices by

$$I_\sigma = \left\{(j_\sigma, l) : \operatorname{supp} u_{(j_\sigma,l)} \subseteq [0,1]\right\}. \tag{3.5}$$

Note that $N_\sigma \asymp 2^{j_\sigma}$. Since $\operatorname{supp} Af \subseteq \operatorname{supp} f$, it follows that for any l, the number of indices l' such that $\operatorname{supp} Au_{j,l} \cap \operatorname{supp} Au_{j,l'} \neq \emptyset$ is bounded by L. Thus, the number of indices l' such that $\langle Au_{j,l}, Au_{j,l'}\rangle_{\mathcal{Y}} > 0$ is also bounded by L. This means that $N_\sigma^* \geq N_\sigma/L$ and $M_\sigma = L$. Consequently, the conditions of Theorem 3.3 are satisfied, and it follows that, in this case, $\mu_\sigma^* \asymp \sqrt{2\sigma^2 \log N_\sigma}$.

Periodic convolution

Let $h : \mathbb{R} \to \mathbb{C}$ be a 1-periodic and bounded function, and let A be the integral operator $A : L^2([0,1]) \to L^2([0,1])$ given by

$$(Af)(x) := \int_0^1 h(u-x)f(u)du, \quad x \in [0,1]. \tag{3.6}$$

The system $(e_k)_{k\in\mathbb{Z}}$, where $e_k(x) = e^{-ikx}$, is a Hilbert basis of $L^2([0,1])$, which consists of singular functions of A, since $A^*Ae_k = |\hat{h}(k)|^2 e_k$. Thus, Corollary 3.4 yields the detection threshold for the detection of anomalies given by $u_k = e_k$.
Let us now try to come up with another system of possible anomalies. For the sake of simplicity, let us, from now on, only consider spaces of real-valued functions, i.e. let $\mathcal{X} = \mathcal{Y} = L^2([0,1],\mathbb{R})$. Motivated by the previous example, let $\{\psi_{j,l} : j,l \in \mathbb{Z}\}$ be a system of compactly supported wavelets with one vanishing moment (i.e. $\int_{\mathbb{R}} x\psi_{j,l}(x)dx = 0$) forming an orthonormal frame of $L^2(\mathbb{R})$. We define periodic wavelets $\psi_{j,l}^{(per)} = \sum_{z\in\mathbb{Z}} \psi_{j,l}(\cdot + z)$ for $l = 0,\dots,2^j - 1$. The system $(u_{(j,l)})_{(j,l)\in I}$ given by $u_{(j,l)} = \psi_{j,l}^{(per)}$ for $I = \bigcup_{j\in\mathbb{Z}}\{j\} \times \{0,\dots,2^j-1\}$ then forms an orthonormal frame of $L^2([0,1])$. If the function h is suffieciently smooth, this constitutes a setting in which, for certain choices of I_σ, Theorem 3.2 cannot be applied and the upper bound from Theorem 3.1 is basically useless, as can be seen in the following lemma.

Lemma 3.7. *Suppose that $h : \mathbb{R} \to \mathbb{R}$ is a 1-periodic, symmetric and continuously differentiable function, and suppose that its derivative h' is Lipschitz. Let $\mathcal{X} = \mathcal{Y} = L^2([0,1],\mathbb{R})$ and let A be the convolution operator defined by (3.6). Let $I = \bigcup_{j\in\mathbb{Z}}\{j\} \times \{0,\dots,2^j-1\} \subseteq \mathbb{R}^2$ and define $u_{(j,l)} = \psi_{j,l}^{(per)}$ as above for any $(j,l) \in I$. Let $(j_\sigma)_{\sigma>0}$ be a family of integers that satisfies $j_\sigma \to \infty$ as $\sigma \to 0$ and set*

$$I_\sigma = \left\{(j_\sigma, l) : l = 0,\dots,2^{j_\sigma} - 1\right\}.$$

for $\sigma > 0$. Then $\gamma_\sigma \to 0$ if $\mu_\sigma/\sigma \to \infty$.

Intuitively, this is explained as follows. When scaled properly, the convolution of a smooth function h with a wavelet with one vanishing moment on a small scale (i.e. when j_σ is large) approximates the derivative of h, but shifted according to the shift parameter of the wavelet (cf. equation (6.15) of [52]). This means that, although the support of $u_{(j_\sigma,l)}$ gets smaller when $\sigma \to 0$, the same is not true for $Au_{(j_\sigma,l)}$. In fact, it turns out, that two possible signals $Au_{(j_\sigma,l)}\|Au_{(j_\sigma,l)}\|_{\mathcal{Y}}^{-1}$ and $Au_{(j_\sigma,l')}\|Au_{(j_\sigma,l')}\|_{\mathcal{Y}}^{-1}$ will be very close (w.r.t. $\|\cdot\|_{\mathcal{Y}}$) when l and l' are close, and hence a test which scans over way less k than in I_σ performs comparably well as (3.2).

Radon transform

Let us finally discuss the example of computerized tomography already mentioned in the introduction. Here, we restrict ourselves to spatial dimension 2, in order to ease readability. We stress, however, that all subsequent results can be extended to any dimensions. Mathematically, this is modeled by the integral operator $R : L^2(\mathcal{B}) \to L^2(Z, (1-t^2)^{-1/2})$, where $\mathcal{B} = \{x \in \mathbb{R}^2 : |x| \leq 1\}$ and $Z = [-1, 1] \times [0, 2\pi)$, given by

$$(Rf)(t, \theta) = \int_{\mathbb{R}} f(t\cos\theta + s\sin\theta, t\cos\theta - s\sin\theta)ds,$$

known as the Radon transform. The singular system of R is analytically known (see [55]). Let $I = \{(k, l) : k \in \mathbb{N}_0,\ |l| \leq k,\ k + l \text{ even}\}$. We define functions $u_{(k,l)} \in L^2(\mathcal{B})$, $(k, l) \in I$ by

$$u_{(k,l)}(x) := e^{il\varphi} r^{|l|} P_k^{(0,|l|)}(2r^2 - 1), \quad x = (r\cos\varphi, r\sin\varphi) \in \mathcal{B},$$

where $P_k^{(0,|l|)}$ are the Jacobi polynomials uniquely determined by the equations $\int_0^1 t^l P_k^{(0,|l|)} P_{k'}^{(0,|l|)} = \delta_{k,k'}$. The system $(u_{(k,l)})_{(k,l)\in I}$ is an orthonormal basis of $L^2(\mathcal{B})$ and, together with the appropriate basis $(v_{(k,l)})_{(k,l)\in I}$ and constants $(\lambda_{(k,l)})_{(k,l)\in I}$ forms the SVD of the Radon transform $R : L^2(\mathcal{B}) \to L^2(Z, (1-t^2)^{-1/2})$. Thus, Corollary 3.4 yields the detection thresholds for the system $(u_{(k,l)})_{(k,l)\in I}$. However, the discussion in Section 3.1 gives rise to another option to choose systems of anomalies that attain the same detection boundaries. For $n \in \mathbb{N}$ we define the usual Sobolev space

$$H^\alpha(\mathbb{R}^n) = \left\{ f \in L^2(\mathbb{R}^n) : \|f\|_{H^\alpha(\mathbb{R}^n)} < \infty \right\},$$

where $\|f\|^2_{H^\alpha(\mathbb{R}^n)} := \int_{\mathbb{R}^2} (1 + |w|)^2)^\alpha |\hat{f}(w)|^2 dw$, and set (in the notation of [55])

$$H_0^\alpha(\mathcal{B}^\circ) := \{ f \in H^\alpha(\mathbb{R}^n) : \operatorname{supp} f \subseteq \mathcal{B} \}.$$

In addition, let

$$H^\alpha(\mathbb{R} \times [0, 2\pi)) := \left\{ f \in L^2(\mathbb{R} \times [0, 2\pi)) : \|f\|_{H^\alpha(\mathbb{R}\times[0,2\pi))} < \infty \right\},$$

where

$$\|f\|^2_{H^\alpha(\mathbb{R}\times[0,2\pi))} := \int_0^{2\pi} \|f(\cdot, \theta)\|^2_{H^\alpha(\mathbb{R})} d\theta.$$

The Radon transform is an operator from $H_0^\alpha(\mathcal{B}^\circ)$ to $H^\alpha(\mathbb{R} \times [0, 2\pi))$ that satisfies (see Theorem 5.1 of [55])

$$C_1 \|f\|_{H_0^\alpha(\mathcal{B}^\circ)} \leq \|Rf\|_{H^\alpha(\mathbb{R}\times[0,2\pi))} \leq C_2 \|f\|_{H_0^\alpha(\mathcal{B}^\circ)}$$

for any $f \in H_0^\alpha(\mathcal{B}^\circ)$. Thus, Theorem 3.6 can be applied. The range of R in $H^\alpha(\mathbb{R} \times [0, 2\pi))$ is $\operatorname{ran}(R) = \{f \in H^\alpha(\mathbb{R} \times [0, 2\pi)) : \operatorname{supp} f \subseteq (-1, 1) \times [0, 2\pi)\}$. Thus, any orthonormal frame $(v_k)_{k\in I}$ of $\operatorname{ran}(R)$ gives rise to a frame $(u_k)_{k\in I}$ of $\mathcal{X} = H_0^\alpha(\mathcal{B}^\circ)$ with sharp detection boundaries given by Theorem 3.6.

3.2 Alternatives given by the linear span of collections of anomalies

Assume now that possibles anomalies might be linear combinations of the u_k, $k \in I_\sigma$. For the upcoming analysis it is necessary to assume that the u_k satisfy the following.

Assumption 3.8. *There is a collection $(v_k)_{k\in I}$ of functions in $\mathcal{Y}$, and a sequence $(\lambda_k)_{k\in I}$ of non-zero complex numbers, such that for any $f \in \mathcal{X}$ it holds that*

$$\langle Af, v_k \rangle_{\mathcal{Y}} = \lambda_k \langle f, u_k \rangle_X.$$

Assumption 3.8 guarantees that we can present our results in terms of the u_k. Clearly, it is satisfied, when $u_k \in \mathrm{ran}(A^*)$ for all $k \in I$. In addition, if we were to assume that the collections $(u_k)_{k\in I}$ and $(v_k)_{k\in I}$ have some kind of useful structure (we may for example assume that they constitute frames of $\mathcal{X}$ and $\mathcal{Y}$, respectively, as we did in Subsection 3.1), then the sequence $(\lambda_k)_{k\in I}$ from Assumption 3.8 takes the role of what might be called quasi-singular values.
In this section, we suppose that $\mathcal{F}_\sigma$ consists of functions in the linear span of the functions u_k, $k \in I_\sigma$, namely $\mathcal{F}_\sigma = \mathcal{F}_\sigma^L = \big\{f \in \mathrm{span}\{u_k : k \in I_\sigma\} : \sum_{k\in I_\sigma} |\lambda_k \langle f, u_k\rangle_{\mathcal{X}}|^2 = 1\big\}$. Thus, testing problem 1.2 becomes

$$H_0 : f = 0 \quad \text{against} \quad H_{1,\sigma} : f \in \mathcal{F}_\sigma^L(\nu_\sigma), \tag{3.7}$$

where

$$\mathcal{F}_\sigma^L(\nu_\sigma) = \left\{ f \in \mathrm{span}\{u_k : k \in I_\sigma\} : \sum_{k\in I_\sigma} |\lambda_k \langle f, u_k\rangle_{\mathcal{X}}|^2 \geq \nu_\sigma^2 \right\}.$$

for some family of positive real numbers $(\nu_\sigma)_{\sigma>0}$ (we use the notation ν_σ instead of μ_σ to avoid confusion with the results from the previous section).

Nonasymptotic results

For a subset $J \subseteq I$, we define the matrix Ξ_J by $(\Xi_J)_{k,k'} = \langle v_k, v_{k'}\rangle_{\mathcal{Y}}$, $k, k' \in J$, and the matrix $\tilde{\Xi}_J$ by $(\tilde{\Xi}_J)_{k,k'} = \langle \tilde{v}_k, \tilde{v}_{k'}\rangle_{\mathcal{Y}}$, $k, k' \in J$, where

$$\tilde{v}_k := \lambda_k^{-1} A u_k,$$

for $k \in I$. We denote the Frobenius norm of a matrix M by $\|M\|_F$.
The next theorem (the non-asymptotic upper bound for the detection threshold) can not be given in terms of the minimax sum of errors γ_σ. Instead we define

$$\gamma_{\sigma,\alpha}(\nu_\sigma) = \inf_{\phi\in\Phi_{\sigma,\alpha}} \left[\alpha_\sigma(\phi) + \sup_{f\in\mathcal{F}_\sigma^L(\nu_\sigma)} \beta_\sigma(\phi, f) \right],$$

where $\Phi_{\sigma,\alpha}$ is the set of all level α tests for the testing problem "H_0 vs $H_{1,\sigma}$". In other words, we consider the minimax sum of errors when only level α tests are allowed.

Theorem 3.9. *Suppose that Assumption 3.8 holds. Assume that the family of subsets $(I_\sigma)_{\sigma>0}$ is such that the matrices Ξ_{I_σ} are positive definite for all $\sigma > 0$. Then, for any $\alpha \in (0,1)$ and $\delta \in (\alpha, 1)$, we have $\gamma_{\sigma,\alpha}(\nu_\sigma) \leq \delta$ if*

$$\nu_\sigma \geq \varepsilon d_\alpha(\delta)\sigma\sqrt{\|\Xi_{I_\sigma}\|_F},$$

where $d_\alpha(\delta) = \sqrt{\log\frac{1}{\delta-\alpha}} + \left(\log\frac{1}{\alpha(\delta-\alpha)} + \sqrt{2\log\frac{1}{\delta-\alpha}} + \sqrt{2\log\frac{1}{\alpha}}\right)^{1/2}$, and ε is given by $\varepsilon = 1$ if $\mathcal{X}$ and $\mathcal{Y}$ are real Hilbert spaces and $\varepsilon = \sqrt{2}$ if $\mathcal{X}$ and $\mathcal{Y}$ are complex Hilbert spaces.

It is now obvious why it is necessary to allow only tests at a prescribed level α. Making α arbitrarily small would require the detection threshold to become arbitrarily large in order to keep the type II error small.
Contrary to the upper bound, the non-asymptotic lower bound for the detection threshold can be stated in terms of γ_σ.

Theorem 3.10. *Suppose that Assumption 3.8 holds, and assume that the family of subsets $(I_\sigma)_{\sigma>0}$ is such that the matrices $\tilde{\Xi}_{I_\sigma}$ are positive definite for all $\sigma > 0$. Then, for any $\delta \in (0,1)$, we have $\gamma_\sigma(\nu_\sigma) \geq \delta$ if*

$$\nu_\sigma \leq c(\delta)\sigma\sqrt{\|\tilde{\Xi}_{I_\sigma}^{-1}\|_F},$$

where $c(\delta) = \big(\log(1 + (2-2\delta)^2)\big)^{1/4}$.

Remark 1: The assumption that Ξ_{I_σ} and $\tilde{\Xi}_{I_\sigma}$, respectively, are positive definite (and consequently invertible, since they are Hermitian) is a technical necessity. However, it is also intuitively justified, because it prevents certain "unreasonable" choices of I_σ (for example any subset I_σ such that $(u_k)_{k\in I_\sigma}$ is linearly dependent).

Remark 2: Note that it can be easily seen that, if we consider the set

$$\left\{ f \in \operatorname{span}\{u_k : k \in I_\sigma\} : \sum_{k\in I_\sigma} |\langle f, u_k\rangle_{\mathcal{X}}|^2 \geq \nu_\sigma^2 \right\},$$

instead of $\mathcal{F}_\sigma^L(\nu_\sigma)$, then we would obtain the same bounds as above with Ξ_{I_σ} replaced by the matrix Λ_{I_σ}, which is given by $(\Lambda_{I_\sigma})_{k,k'} = (\lambda_k\overline{\lambda_{k'}})^{-1}\langle v_k, v_{k'}\rangle_{\mathcal{Y}}$, and $\tilde{\Xi}_{I_\sigma}$ replaced by the matrix $\tilde{\Lambda}_{I_\sigma}$ given by $(\tilde{\Lambda}_{I_\sigma})_{k,k'} = \langle Au_k, Au_{k'}\rangle_{\mathcal{Y}}$ It follows, that our results are compatible with the results obtained in [49], where the above testing problem was considered when the system $(u_k, v_k, \lambda_k)_{k\in I}$ is given by the SVD of A.

Asymptotic results

The asymptotic results for this section can now be easily deduced from the previous theorems.

Corollary 3.11. *Suppose that the assumptions of Theorems 3.9 and 3.10 hold.*

(1) H_0 and $H_{1,\sigma}$ are asymptotically distinguishable if

$$\frac{\nu_\sigma}{\sigma\sqrt{\|\Xi_{I_\sigma}\|_F}} \to \infty.$$

(2) H_0 and $H_{1,\sigma}$ are asymptotically indistinguishable if

$$\frac{\nu_\sigma}{\sigma\sqrt{\|\tilde{\Xi}_{I_\sigma}^{-1}\|_F}} \to 0.$$

Until now, we have allowed the u_k to just be any functions we might be interested in detecting. However, we are able to refine our results when we assume that $(v_k)_{k\in I}$ and $(\tilde{v}_k)_{k\in I}$ are "well-behaved". We call a sequence of functions $(h_i)_{i\in\mathbb{N}}$ from some Hilbert space $\mathcal{H}$ a *Riesz sequence*, if there exist constants $C_1, C_2 > 0$ such that

$$C_1 \sum_{i\in\mathbb{N}} |a_i|^2 \leq \| \sum_{i\in\mathbb{N}} a_i h_i\|_{\mathcal{H}}^2 \leq C_2 \sum_{i\in\mathbb{N}} |a_i|^2,$$

for any sequence $(a_i)_{i\in\mathbb{N}} \in \ell^2$. Two sequences $(h_i)_{i\in\mathbb{N}}$ and $(h_i')_{i\in\mathbb{N}}$ are called *biorthogonal* if

$$\langle h_i, h_j'\rangle_{\mathcal{H}} = \delta_{i,j},$$

where $\delta_{i,j}$ is the Kronecker symbol.

Assumption 3.12. *The collections $(v_k)_{k\in I}$ and $(\tilde{v}_k)_{k\in I}$ are biorthogonal Riesz sequences.*

We acknowledge that Assumption 3.12 is restrictive. We will discuss non-trivial situations in which it is satisfied below. We collect the implications of Assumption 3.12 in the following lemma.

Lemma 3.13. *Suppose that Assumptions 3.8 and 3.12 hold, and let $(I_\sigma)_{\sigma>0}$ be an arbitrary family of subsets of I. Then the following statements hold.*

(1) For any $\sigma > 0$, the matrices Ξ_{I_σ} and $\tilde{\Xi}_{I_\sigma}$ are positive definite.

(2) There are constants $c_1, c_2 > 0$ such that $c_1\|\Xi_{I_\sigma}\|_F \leq \|\tilde{\Xi}_{I_\sigma}^{-1}\|_F \leq c_2\|\Xi_{I_\sigma}\|_F$.

(3) $\|\Xi_{I_\sigma}\|_F \sim N_\sigma^{1/2}$ (and consequently, also $\|\tilde{\Xi}_{I_\sigma}^{-1}\|_F \sim N_\sigma^{1/2}$), as $\sigma \to 0$, where $N_\sigma = |I_\sigma|$.

Thus, if all conditions of Lemma 3.13 are satisfied, it follows from Corollary 3.11 that the *separation rate* of the family of testing problems "H_0 against $H_{1,\sigma}$" is given by $\nu_\sigma^* \sim \sigma N_\sigma^{1/4}$.

Remark: Note that this result is not surprising. The separation rate corresponds to the rate (in terms of the euclidean norm) of detecting an n-dimensional signal $\theta \in \mathbb{R}^n \setminus \{0\}$ from observations given by $X = \theta + \sigma Z$, where $Z \sim \mathcal{N}(0, \mathrm{id}_n)$. Furthermore, it has been shown previously (cf. [49]) that the same holds, when $(u_k, v_k, \lambda_k)_{k\in I}$ constitute the SVD of the operator A. Thus, the above results yield a generalization of the known theory.

Examples

It is clear that, when the system $(u_k, v_k, \lambda_k)_{k\in I}$ is given by the SVD of the operator $A : \mathcal{X} \to \mathcal{Y}$, then all of the above theory can be applied. Since this was the subject of [49], we will omit a discussion of this example here.

Examples based on the wavelet-vaguelette decomposition

Suppose that $(u_k)_{k\in I}$ is a system of orthogonal wavelets in $\mathcal{X}$. If chosen appropriately (for a complete discussion, see [18]), it follows that for certain operators $A : \mathcal{X} \to \mathcal{Y}$, there exist non-zero numbers $(\lambda_k)_{k\in I}$, such that the systems $(v_k)_{k\in I}$ and $(\tilde{v}_k)_{k\in I}$ of functions in $\mathcal{Y}$ given by

$$A^* v_k = \lambda_k u_k, \quad \tilde{v}_k = \lambda_k^{-1} A u_k$$

form biorthogonal Riesz sequences in $\mathcal{Y}$ (see Theorem 2 of [18])). Clearly, Assumption 3.8 is satisfied in this case.
We immediately see that this would yield nice examples of the theory developed in this section. We will discuss a few situations, in which such a construction is possible, below.

Integration

Consider the setting of example 3.1, i.e. $u_{(j,l)} = \psi_{j,k}$ for $(j,l) \in I = \mathbb{Z}^2$ and for some wavelet ψ. Suppose that the wavelet ψ is continuously differentiable. In this case, the WVD is particularly simple. Let

$$v_{(j,l)}(x) = -2^{j/2}\psi'(2^j x - l), \quad \tilde{v}_{(j,l)}(x) = 2^{j/2}\psi^{(-1)}(2^j x - l),$$

with $\lambda_{(j,l)} = 2^{-j}$. Then it follows from [18] that the systems $(v_{(j,l)})_{(j,l)\in I}$ and $(\tilde{v}_{(j,l)})_{(j,l)\in I}$ form biorthogonal Riesz sequences in $L^2(\mathbb{R})$. Thus, we can apply Lemma 3.13 to obtain $\|\Xi_{I_\sigma}\|_F \sim N_\sigma^{1/2}$, and thus, $\nu_\sigma^* \sim \sigma N_\sigma^{1/4}$ for any family $(I_\sigma)_{\sigma>0}$ of "candidate" indices.

Periodic Convolution

Let $\mathcal{X} = \mathcal{Y} = L^2(\mathcal{S}^1)$, where $\mathcal{S}^1$ is the unit circle. In other words we consider square-integrable 1-periodic functions on $[0,1]$. Let the operator A be given by $(Af)(x) = \int_0^1 h(x-u)f(u)du$, $x \in [0,1]$ for some 1-periodic function h. Let $(u_{j,k})_{(j,k)\in\mathbb{Z}^2}$ be a basis of periodic Meyer wavelets, each with Fourier coefficients $u_{j,k,m}$, $m \in \mathbb{Z}$, and let $(I_\sigma)_{\sigma>0}$ be any family of finite subsets of $I = \mathbb{Z}^2$. Let h_m, $m \in \mathbb{Z}$ be the Fourier coefficients of h. It was shown in Appendix B of [43] that, if $h_m = C|m|^{-a}$, for some $a > 0$, the collections $(v_{j,k})_{(j,k)\in\mathbb{Z}^2}$ and $(\tilde{v}_{j,k})_{(j,k)\in\mathbb{Z}^2}$ given by

$$v_{j,k}(x) = \sum_{m\in\mathbb{Z}} \frac{\lambda_{j,k} u_{j,k,m}}{h_m} e^{imx}, \quad \tilde{v}_{j,k}(x) = \sum_{m\in\mathbb{Z}} \frac{h_m u_{j,k,m}}{\lambda_{j,k}} e^{imx},$$

where the quasi-singular values are given by $\lambda_{j,k} = 2^{-ja}C$, yield biorthogonal Riesz sequences.

Radon transform

We start by introducing two-dimensional wavelet systems. Let ψ be a (one-dimensional) wavelet and φ its corresponding scaling function. We assume that they are of compact support and at least two times continuously differentiable. Define the two-dimensional functions

$$\eta^0(x) := \varphi(x_1)\theta(x_2), \quad \eta^1(x) := \psi(x_1)\varphi(x_2), \quad \eta^2(x) := \varphi(x_1)\psi(x_2), \quad \eta^3(x) := \psi(x_1)\psi(x_2),$$

where $x = (x_1, x_2) \in \mathbb{R}^2$. The function η^0 is the two-dimensional scaling function, and the functions η^ε, $\varepsilon \in \{1,2,3\}$, are the two-dimensional wavelets. They are scaled and translated as usual, i.e.

$$\eta^\varepsilon_{j,l}(x) := 2^j \eta^\varepsilon(2^j x - l), \quad j \in \mathbb{Z},\ l \in \mathbb{Z}^2,\ \varepsilon \in \{0,1,2,3\}.$$

The collection of all translated and scaled wavelets (without the scaling functions, i.e. excluding $\varepsilon = 0$) is a Hilbert basis of $L^2(\mathbb{R}^2)$ (see for example Theorem 7.24 of [52]). Using the projection theorem (see Theorem 1.1 of [55]), it was shown in [18] that for $\varepsilon \in \{1,2,3\}$ and for $f \in L^2(\mathbb{R}^2) \cap \mathrm{dom}(R)$, we have

$$\int_0^\pi \int_{-\infty}^\infty (Rf)(t,\theta)\overline{(R\omega^\varepsilon_{j,k})(t,\theta)}dtd\theta = \int_{\mathbb{R}^2} f(x)\overline{\eta^\varepsilon_{j,k}(x)}dx$$

where $\omega^\varepsilon_{j,k}$ is defined through its Fourier transform $\widehat{\omega^\varepsilon_{j,k}}(x) = \frac{1}{2\pi}|x|\widehat{\eta^\varepsilon_{j,k}}(x)$. In other words,

$$\langle Rf, R\omega^\varepsilon_{j,k}\rangle_{L^2(\mathbb{R}\times[0,2\pi))} = \langle f, \eta^\varepsilon_{j,k}\rangle_{L^2(\mathbb{R}^2)}. \tag{3.8}$$

An in-homogeneous wavelet basis of $L^2(\mathbb{R})$ can be constructed as follows: For some $j_0 \in \mathbb{Z}$, we consider the collection of functions

$$\left\{\eta^0_{j_0,k} : k \in \mathbb{Z}^2\right\} \cup \left\{\eta^\varepsilon_{j,k} : \varepsilon \in \{1,2,3\}, k \in \mathbb{Z}^2, j \geq j_0\right\}.$$

Note that for any $j_0 \in \mathbb{Z}$, $k_0 \in \mathbb{Z}^2$ we can write

$$\eta^0_{j_0,k_0} = \sum_{\varepsilon\in\{1,2,3\}} \sum_{k\in\mathbb{Z}^2} \sum_{-\infty<j\leq j_0} c^\varepsilon_{j_0,k_0,j,k}\eta^\varepsilon_{j,k},$$

and thus, it follows from the linearity of all the above operations that the relation (3.8) is also true for $\varepsilon = 0$ with $w^0_{j,k}$ defined accordingly.

With practical applications (where it is an unreasonable assumption that observations on all of the plane $\mathbb{R}^2$ can be made) in mind, we assume that signals, if existent, lie within a compact set, e.g. the unit ball $\mathcal{B} = \{x \in \mathbb{R}^2 : \|x\|_2 \leq 1\}$. We consider the Radon transform as an operator $R : \mathcal{X} \to \mathcal{Y}$, where $\mathcal{X} = L^2(B)$, and $\mathcal{Y} = L^2(Z)$ with $Z = [-1,1] \times [0,\pi)$. Note that, contrary to the example in Section 3.1, the space $\mathcal{Y}$ is equipped with the norm $\|\cdot\|_\mathcal{Y}$ given by $\|f\|_\mathcal{Y} = \int_{-1}^1 \int_0^{2\pi} |f(t,\theta)|^2 d\theta dt$. (Note that the operator $\mathbb{R} : \mathcal{X} \to \mathcal{Y}$ is well-defined and bounded since $\|f\|_{L^2(Z)} \leq \|f\|_{L^2(Z,(1-t)^{-1/2})}$ for any $f \in L^2(Z)$.)

We devise a "wavelet-type" frame of $L^2(\mathcal{B})$ as follows. We choose j_0 large enough, such that the area of $\mathrm{supp}(\eta^\varepsilon_{j_0,k})$ is small compared to the area of the unit ball $\mathcal{B}$. Now let

$$I = \left\{(j_0,k,0) : \mathrm{supp}(\eta^0_{j_0,k}) \cap \mathcal{B} \neq \varnothing\right\} \cup \left\{(j,k,\varepsilon) : j \geq j_0,\ \varepsilon \in \{1,2,3\},\ \mathrm{supp}(\eta^\varepsilon_{j,k}) \cap \mathcal{B} \neq \varnothing\right\},$$

and finally, define $u_{(j,k,\varepsilon)} = \eta^\varepsilon_{j,k}|_\mathcal{B}$ for $(j,k,\varepsilon) \in I$. The collection $(u_{(j,k,\varepsilon)})_{(j,k,\varepsilon)\in I}$ forms a frame of $L^2(\mathcal{B})$, since, for any $f \in L^2(\mathbb{R}^2)$ supported in $\mathcal{B}$, we have

$$\langle f, u_{(j,k,\varepsilon)}\rangle_{L^2(\mathcal{B})} = \langle f, \eta^\varepsilon_{j,k}\rangle_{L^2(\mathbb{R}^2)}.$$

Note that $\|R\omega^\varepsilon_{(j,k)}\|_{L^2(\mathbb{R}\times[0,2\pi))} \sim 2^{j/2}$ (again, see [18]). Thus, if we let $v_{(j,k,\varepsilon)} = 2^{-j/2}R\omega^\varepsilon_{(j,k)}|_Z$, we obtain

$$\langle Rf, v_{(j,k,\varepsilon)}\rangle_{L^2(Z)} = 2^{-j/2}\langle Rf, R\omega^\varepsilon_{(j,k)}\rangle_{L^2(\mathbb{R}\times[0,2\pi))} = 2^{-j/2}\langle f, \eta^\varepsilon_{(j,k)}\rangle_{L^2(\mathbb{R}^2)} = 2^{-j/2}\langle f, u_{(j,k,\varepsilon)}\rangle_{L^2(\mathcal{B})},$$

for any $f \in L^2(\mathcal{B})$ (which we extended to $L^2(\mathbb{R}^2)$ by setting $f(x) = 0$ whenever $x \notin \mathcal{B}$). Thus, Assumption 3.8 is satisfied for the set $(u_{(j,k,\varepsilon)})_{(j,k,\varepsilon)\in I}$ with $v_{(j,k,\varepsilon)}$ defined as above and $\lambda_{(j,k,\varepsilon)} = 2^{-j/2}$.

It follows that Theorems 3.9 and 3.10 are applicable for the collection $(u_{(j,k,\varepsilon)})_{(j,k,\varepsilon)\in I}$ and yield non-asymptotic results for appropriate choices of I_σ. However, note that the $u_{(j,k,\varepsilon)}$ are not necessarily orthonormal.

It follows from Lemma 4 (and the discussion leading up to it) of [18] that, if ψ has at least 4 vanishing moments and is at least 4 times continuously differentiable, then the collections $(2^{-j/2}R\omega^{\varepsilon}_{j,k})_{j\in\mathbb{Z},k\in\mathbb{Z}^2,\varepsilon\in\{1,2,3\}}$ and $(2^{j/2}R\eta^{\varepsilon}_{j,k})_{j\in\mathbb{Z},k\in\mathbb{Z}^2,\varepsilon\in\{1,2,3\}}$ are Riesz sequences.
If we suppose that all subsets I_σ are chosen such that all $\eta^{\varepsilon}_{j,k}$ lie completely within $\mathcal{B}$, i.e. $\operatorname{supp}(\eta^{\varepsilon}_{j,k}) \subseteq \mathcal{B}$ for any $(j,k,\varepsilon) \in I_\sigma$ for all $\sigma > 0$, then $\tilde{v}_{(j,k,\varepsilon)} = 2^{j/2}R\eta^{\varepsilon}_{j,k}$, and it follows from the same arguments as in the proof of Lemma 3.13 that

$$\|\tilde{\Xi}^{-1}_{I_\sigma}\|_F \geq C\|\Omega_{I_\sigma}\|_F \geq C'\sqrt{N_\sigma},$$

where Ω_{I_σ} is given by $(\Omega_{I_\sigma})_{k,k'} = \langle 2^{-j/2}R\omega^{\varepsilon}_{j,k}, 2^{-j/2}R\omega^{\varepsilon}_{j,k'}\rangle_{L^2(\mathbb{R}\times[0,2\pi))}$. On the other hand, since $(v_k)_{k\in I}$ is a frame of $L^2([-1,1]\times[0,\pi))$, it follows (as in the proof of Lemma 3.13) that

$$\|\Xi_{I_\sigma}\|^2_F \leq C'' \sum_{k\in I_\sigma} \|v_k\|_{L^2([-1,1]\times[0,\pi)} \leq C'' \sum_{k\in I_\sigma} \|2^{-j/2}R\omega^{\varepsilon}_{j,k}\|_{L^2(\mathbb{R}\times[0,\pi)} \leq C'''N_\sigma.$$

Thus, $\nu^*_\sigma \sim \sigma N_\sigma^{1/4}$.

4 Simulation study

A note on discretization

For convenience, we will from now on assume that $\mathbb{K} = \mathbb{R}$. Assume that $\mathcal{Y} = L^2(D,\mathbb{R})$ for some $D \subseteq \mathbb{R}^d$, $d \in \mathbb{N}$. For a finite subset $S \subseteq D$, we define the evaluation function e_S by $e_S : h \mapsto (h(s))_{s\in S}$. Now let $A_S = e_S \circ A$. Clearly, $A_S : \mathcal{X} \to \mathbb{R}^n$, where $n = |S|$, is a bounded linear operator. We equip $\mathbb{R}^n$ with the inner product $\langle\cdot,\cdot\rangle_n$ (and the corresponding norm $\|\cdot\|_n$) given by $\langle x, x'\rangle_n^2 = \frac{\operatorname{vol}(D)}{n}\sum_{s\in S} x_s x'_s$, for $x = (x_s)_{s\in S}, x' = (x'_s)_{s\in S} \in \mathbb{R}^n$, and thereby make $\mathbb{R}^n$ a Hilbert space. Here, $\operatorname{vol}(D)$ denotes the (d-dimensional) volume of D. Now suppose that we observe data $Y_{\sigma,S}$ on $\mathbb{R}^n$ given by

$$Y_{\sigma,S} = A_S(f) + \sigma\sqrt{\frac{n}{\operatorname{vol}(D)}}\xi, \tag{4.1}$$

where $\xi = (\xi_s)_{s\in S} \sim \mathcal{N}(0, \mathrm{id}_n)$. Since, for any $x, x' \in \mathbb{R}^n$, we have

$$\left\langle\sqrt{\frac{n}{\operatorname{vol}(D)}}\xi, x\right\rangle_n \sim \mathcal{N}(0, \|x\|_n) \quad \text{and} \quad \mathbb{E}\left(\left\langle\sqrt{\frac{n}{\operatorname{vol}(D)}}\xi, x\right\rangle_n \left\langle\sqrt{\frac{n}{\operatorname{vol}(D)}}\xi, x'\right\rangle_n\right) = \langle x, x'\rangle_n,$$

it follows that our results are valid for this discretized model. Note that asymptotic results still refer to σ becoming small (and not n becoming large). If S is chosen appropriately, $\langle e_S(h), e_S(h')\rangle_n$ can be viewed as an approximation of $\langle h, h'\rangle_{\mathcal{Y}}$ for $h, h' \in \mathcal{Y}$, and, consequently, the upper and lower bounds derived from (4.1) can be viewed as an approximation of the upper and lower bounds derived from the data (1.1).
Finally, note that testing "$f = 0$" against "$f \in \mathcal{F}_\sigma(\mu_\sigma)$" based on the data $Y_{\sigma,n}$ is equivalent to testing "$f = 0$" against "$f \in \mathcal{F}_\sigma\left(\frac{\mu_\sigma\operatorname{vol}(D)}{n}\right)$" based on $X_{\sigma,S}$ given by

$$X_{\sigma,S} := A_S(f) + \sigma\sqrt{\frac{\operatorname{vol}(D)}{n}}\xi.$$

Integration

We consider the example from section 3.1, discretized as above with $S = \{\frac{i}{n} : i = 0,\dots,n-1\}$ for $n = 2^{15}$. The wavelet system $(\psi_{j,k})_{j,k\in\mathbb{Z}}$ consists of Daubechies (db6) wavelets. See Figure 2 for the results of the simulation study. Note that the displayed results are only approximations in two senses: First, we used the test $\phi_{\sigma,\alpha} = \phi^{ML}_{\sigma,\alpha}$ from (3.2), which may not necessarily be the optimal test, and second, we approximate $\sup_{f\in\mathcal{F}_\sigma(\delta)}\beta_\sigma(\phi^{ML}_{\sigma,\alpha}, f)$ by $\beta^*_\sigma(\delta) :=$

$N_\sigma^{-1} \sum_{k \in I_\sigma} \beta_\sigma(\phi^{ML}_{\sigma,\alpha}, \delta \|Au_k\|_{\mathcal{Y}}^{-1} u_k)$, i.e. the mean type II error over all possible anomalies of minimal "amplitude", since it is, in general, not clear, which $k \in I_\sigma$ will maximize the type II error.
Note that the proof of Theorem 3.1 yields a non-asymptotic upper bound for the detection threshold: We have $\sup_{f \in \mathcal{F}_\sigma(\delta)} \beta_\sigma(\phi^{ML}_{\sigma,\alpha}, f) \leq \alpha$ when $\delta \geq \sigma(c_{\alpha,\sigma} + z_{1-\alpha})$, where $z_{1-\alpha}$ is the $(1-\alpha)$-quantile of the standard Gaussian distribution.
Next, we consider the example from section 3.2. Everything is as above, except for a few differences: We consider alternatives given by linear combinations of $(u_k)_{k \in I_\sigma}$ as in Section 3.2, we use the test $\phi_{\sigma,\alpha} = \phi^{\chi^2}_{\sigma,\alpha}$ given by (6.2), and $\beta^*_\sigma(\delta)$ is given by $\mathbb{E}_\pi \beta_\sigma(\phi^{ML}_{\sigma,\alpha}, \delta f)$, where π is the uniform distribution on $\mathcal{F}^L_\sigma$. The results of this study are displayed in Figure 3.

Radon transform

The setting for our simulation study for the Radon transform is inspired by the discussion in Section 3.2. We consider the Radon transform as an operator

$$R : L^2(\mathcal{B}_{\frac{1}{\sqrt{2}}}, \mathbb{R}) \to L^2\left(\left[-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right] \times [0, \pi), \mathbb{R}\right),$$

where $\mathcal{B}_{\frac{1}{\sqrt{2}}} = \left\{x \in \mathbb{R}^2 : \|x\|_2 \leq \frac{1}{\sqrt{2}}\right\}$ is the ball that contains the unit square $[-1/2, 1/2]^2$. Let $\{\eta^\varepsilon_{j,l} : j \in \mathbb{Z}, l \in \mathbb{Z}^2, \varepsilon \in \{1,2,3\}\}$ be the two-dimensional wavelet system (consisting of Daubechies (db4) wavelets) from Section 3.2, define $u_{(j,l,\varepsilon)} = \eta^\varepsilon_{j,l}$ for $(j,l,\varepsilon) \in I$ with

$$I = \left\{(j,l,\varepsilon) : \operatorname{supp} \eta^\varepsilon_{j,l} \subseteq [-1/2, 1/2]^2\right\},$$

and let $I_\sigma = \left\{(j_\sigma, l, \varepsilon) : \operatorname{supp} \eta^\varepsilon_{j_\sigma,l} \subseteq [-1/2, 1/2]^2\right\}$, for family $(j_\sigma)_{\sigma>0}$ of natural numbers. We consider discretized data of the form (4.1) with $S = \left\{\left(-\frac{1}{2} + \frac{i_1}{1024}, \frac{i_2}{360}\pi\right) : i_1 = 0, \ldots, 1023, i_2 = 0, \ldots, 359\right\}$.
As above, we use the test $\phi_{\sigma,\alpha} = \phi^{\chi^2}_{\sigma,\alpha}$ given by (6.2), and $\beta^*_\sigma(\delta) = \mathbb{E}_\pi \beta_\sigma(\phi^{ML}_{\sigma,\alpha}, \delta f)$, where π is the uniform distribution on $\mathcal{F}^L_\sigma$. The results of this study are displayed in Figure 4.
The example in Figure 1 also comes from this setting: The parameters were $j_\sigma = 5$, $\sigma = 15$, and $\delta = 264$. By Theorem 3.9, the distorted image can be distinguished from the reference image with type I and type II error both at most 0.05 by the test $\phi^{\chi^2}_{8,0.05}$.

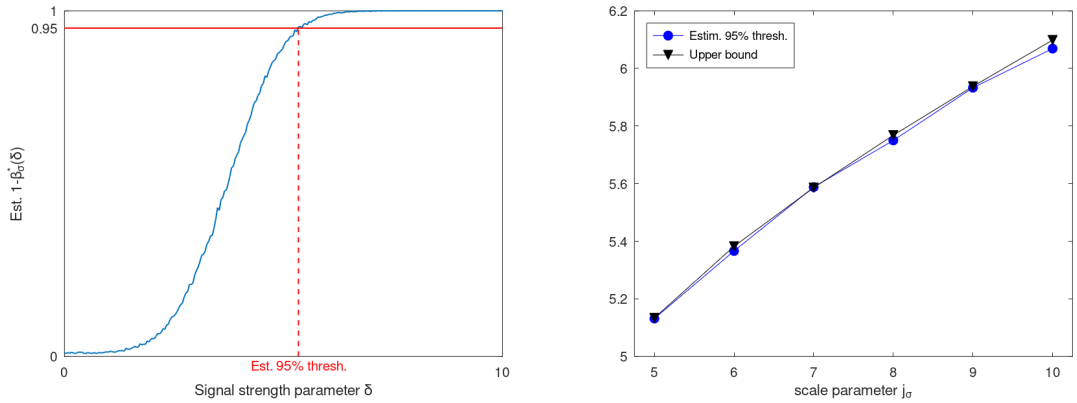


Figure 2: **Left:** Estimation of $1-\beta^*_\sigma(\delta)$ for δ between 0 and 10, for $j_\sigma = 6$, $\sigma = 1$ and $\alpha = 0.05$. For each value of δ, 5000 tests have been performed. The results suggest that the power achieves 95% for $\delta \approx 5.3414$. **Right:** Estimated values of δ for which the power achieves 95% for $j_\sigma \in \{5, \ldots, 10\}$ and $\sigma = 1$ compared with the upper bound derived from the proof of Theorem 3.1.

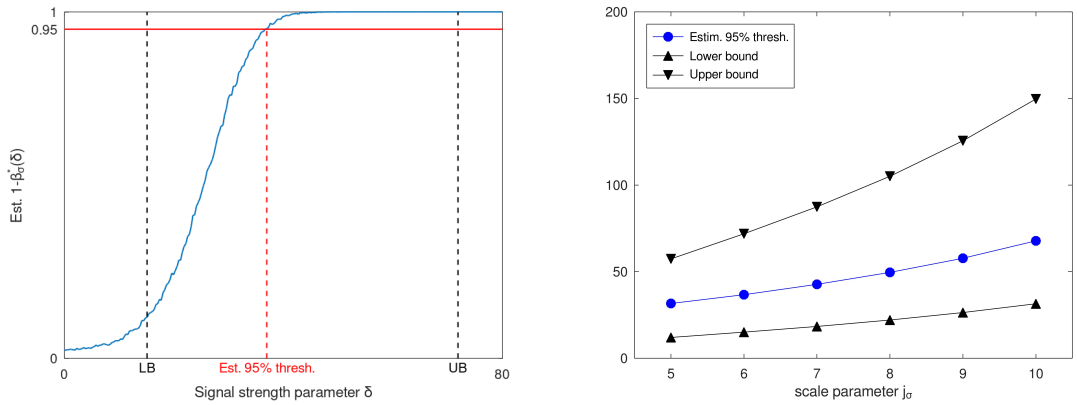


Figure 3: **Left:** Estimation of $1 - \beta^*_\sigma(\delta)$ for δ between 0 and 80, for $j_\sigma = 6$, $\sigma = 1$ and $\alpha = 0.05$, compared with the upper bound (UB) from 3.9 and the lower bound (LB) from Theorem 3.10. For each value of δ, 5000 tests have been performed. The results suggest that the power achieves 95% for $\delta \approx 36.948$. **Right:** Estimated values of δ for which the power achieves 95% for $j_\sigma \in \{5, \dots, 10\}$ and $\sigma = 1$ compared with the upper bound from 3.9 and the lower bound from Theorem 3.10.

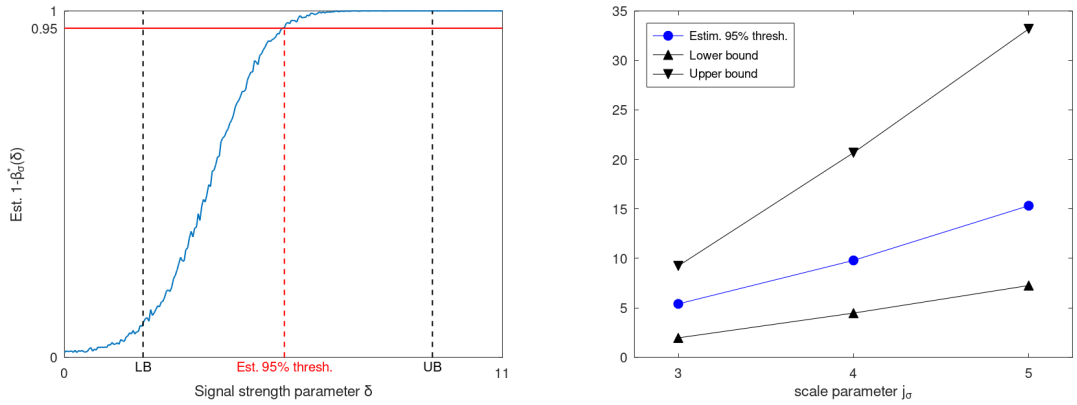


Figure 4: **Left:** Estimation of $1 - \beta^*_\sigma(\delta)$ for δ between 0 and 11, for $j_\sigma = 3$, $\sigma = 1$ and $\alpha = 0.05$, compared with the upper bound (UB) from 3.9 and the lower bound (LB) from Theorem 3.10. For each value of δ, 5000 tests have been performed. The power achieves 95% for $\delta \approx 5.5221$. **Right:** Estimated values of δ for which the power achieves 95% for $j_\sigma \in \{5, \dots, 10\}$ and $\sigma = 1$ compared with the upper bound from 3.9 and the lower bound from Theorem 3.10.

These simulations seem to affirm our theoretical results. Note that the thresholds displayed in Figure 2 are very large compared to the asymptotic detection boundary from Corollary 3.3. This is due to the logarithmic growth of the detection boundary.

5 Discussion

In this paper, we have considered statistical hypothesis testing in inverse problems with localized alternatives. This can be used to determine whether an unknown object that deviates from a reference object, can be distinguished from that reference or not.
More precisely, we first considered alternatives given by finitely many elements (e.g. chosen from a dictionary), and under additional restrictions on the structure of this system we were able to derive the (asymptotic) detection boundary. Those results are illustrated along examples such as

integration, convolution, and the Radon transform. Afterwards, we have moved to more complex alternatives allowing for linear combinations of elements from the dictionary. In this case, we were still able to derive the minimax separation rate even under weaker assumptions on the structure of the system. This has been illustrated again for the above-mentioned and in simulations.
The results in this study offer several point of contact for further research. For practical purposes, the design of more (computationally and statistically) efficient multiple tests is on demand, which is beyond the scope of this paper. It would be interesting to see which methods can be used to efficiently test a reference object against hypotheses which consist e.g. of wavelets on different scales, which is a setting, for which the assumptions of Corollary 3.3 are not satisfied in general. Another interesting question is the detection boundary in case of sparse alternatives (similar to [49]), which we have not discussed here.

Acknowledgements

M. P. acknowledges support from the RTG 2088. A. M. acknowledges support of the DFG Cluster of Excellence 2067 "Multiscale Bioimaging: From Molecular Machines to Networks of Excitable Cells". F. W. is supported by the DFG via grant WE 6204/4-1. The authors would like to thank Markus Haltmeier and Miguel del Álamo for insightful comments.

6 Proofs

6.1 Proof for section 3.1

Proof of the upper bound

Proof of Theorem 3.1. We treat the two cases (whether $\mathcal{X}$ and $\mathcal{Y}$ are real or complex spaces) separately.

$\mathcal{X}$ and $\mathcal{Y}$ are real Hilbert spaces. Any test for the testing problem (1.2) yields an upper bound for γ_σ, and, thus, also an upper bound for μ_σ^*. Our upper bound is based on a particularly simple family of likelihood ratio type tests given by (3.2) with thresholds given by

$$c_{\alpha,\sigma} = \sqrt{2 \log \frac{N_\sigma}{\alpha}}.$$

We show that for any $\sigma > 0$ and any $\alpha \in (0,1)$, the test $\phi_{\alpha,\sigma}$ has level α and its asymptotic type II error vanishes for the testing problem (1.2) if $\mu_\sigma \gtrsim (1+\varepsilon_\sigma)\sqrt{2\sigma^2 \log N_\sigma}$. This would then imply that $\gamma_\sigma \precsim \alpha + o(1)$, which will immediately prove the theorem, since α was arbitrary.
Setting $f_k = \frac{u_{k'}}{\|Au_{k'}\|_\mathcal{Y}}$, we have

$$\frac{\langle Y_\sigma, Au_k\rangle_\mathcal{Y}}{\sigma\|Au_k\|_\mathcal{Y}} \overset{H_0}{\sim} \mathcal{N}(0,1) \quad \text{and} \quad \frac{\langle Y_\sigma, Au_k\rangle_\mathcal{Y}}{\sigma\|Au_k\|_\mathcal{Y}} \overset{f=\delta f_k}{\sim} \mathcal{N}\left(\frac{\delta\langle Au_{k'}, Au_k\rangle_\mathcal{Y}}{\sigma\|Au_{k'}\|_\mathcal{Y}\|Au_k\|_\mathcal{Y}}, 1\right)$$

Using the union bound and a concentration inequality for the normal distribution we find

$$\begin{aligned}\mathbb{P}_{H_0}\left(\phi_{\alpha,\sigma}(y_\sigma) = 1\right) &= \mathbb{P}_{H_0}\left(\sup_{k\in I_\sigma} \frac{|\langle Y_\sigma, Au_k\rangle_\mathcal{Y}|}{\sigma\|Au_k\|_\mathcal{Y}} > c_{\alpha,\sigma}\right) \\ &\leq N_\sigma \mathbb{P}\left(|Z| > c_{\alpha,\sigma}\right) = N_\sigma \exp\left(-\frac{1}{2}c_{\alpha,\sigma}^2\right) = \alpha,\end{aligned}$$

for some $Z \sim \mathcal{N}(0,1)$. Thus, $\phi_{\alpha,\sigma}$ is indeed a level α test. Next, we show that the maximal type II error of $\phi_{\alpha,\sigma}$ vanishes. For some random variables $Z_{\sigma,k} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$, $k \in I_\sigma$, we find

$$\begin{aligned}
\sup_{k\in I_\sigma}\sup_{|\delta|\geq\mu_\sigma} \mathbb{P}_{\delta f_k}\left(\phi_{\alpha,\sigma}(y_\sigma)=0\right) &= \sup_{k\in I_\sigma}\sup_{|\delta|\geq\mu_\sigma} \mathbb{P}_{\delta f_k}\left(\sup_{k'\in I_\sigma}\frac{|\langle Y_\sigma, Au_{k'}\rangle_{\mathcal{Y}}|}{\sigma\|Au_{k'}\|_{\mathcal{Y}}} \leq c_{\alpha,\sigma}\right) \\
&\leq \sup_{k\in I_\sigma}\sup_{|\delta|\geq\mu_\sigma} \mathbb{P}\left(\sup_{k'\in I_\sigma}\left(Z_{\sigma,k'} + \frac{\delta\langle Au_{k'}, Au_k\rangle_{\mathcal{Y}}}{\sigma\|Au_{k'}\|_{\mathcal{Y}}\|Au_k\|_{\mathcal{Y}}}\right) \leq c_{\alpha,\sigma}\right) \\
&\leq \sup_{k\in I_\sigma}\sup_{|\delta|\geq\mu_\sigma} \mathbb{P}\left(Z_{\sigma,k} + \frac{\delta}{\sigma} \leq c_{\alpha,\sigma}\right) \\
&= \mathbb{P}\left(Z + \frac{\mu_\sigma}{\sigma} \leq c_{\alpha,\sigma}\right) \\
&= \mathbb{P}\left(Z \leq \sqrt{2\log N_\sigma} + \sqrt{2\log(1/\alpha)} - (1+\varepsilon_\sigma)\sqrt{2\log N_\sigma}\right) \to 0,
\end{aligned}$$

since $\varepsilon_\sigma\sqrt{\log N_\sigma} \to \infty$.

$\mathcal{X}$ and $\mathcal{Y}$ are complex Hilbert spaces. The idea of the proof is the same as above. We again use is the test given by (3.2) with thresholds

$$c_{\alpha,\sigma} = \sqrt{1 + 2\sqrt{\log\frac{N_\sigma}{\alpha}} + 2\log\frac{N_\sigma}{\alpha}}.$$

Setting $f_{k'} = \frac{u_{k'}}{\|Au_{k'}\|_{\mathcal{Y}}}$ as above, we have

$$\frac{\langle Y_\sigma, Au_k\rangle_{\mathcal{Y}}}{\sigma\|Au_k\|_{\mathcal{Y}}} \overset{H_0}{\sim} \mathcal{CN}(0,2) \quad \text{and} \quad \frac{\langle Y_\sigma, Au_k\rangle_{\mathcal{Y}}}{\sigma\|Au_k\|_{\mathcal{Y}}} \overset{f=\delta f_{k'}}{\sim} \mathcal{CN}\left(\frac{\delta\langle Au_{k'}, Au_k\rangle_{\mathcal{Y}}}{\sigma\|Au_{k'}\|_{\mathcal{Y}}\|Au_k\|_{\mathcal{Y}}}, 2\right)$$

We first show that Φ_α is a level α test. For some $Z \sim \mathcal{CN}(0,1)$ we find, using the union bound, that

$$\mathbb{P}_{H_0}\left(\phi_{\alpha,\sigma}(y_\sigma)=1\right) = \mathbb{P}_{H_0}\left(\sup_{k\in I_\sigma}\frac{|\langle Y_\sigma, Au_k\rangle_{\mathcal{Y}}|}{\sigma\|Au_k\|_{\mathcal{Y}}} > c_{\alpha,\sigma}\right) \leq N_\sigma\mathbb{P}\left(|Z|^2 > c_{\alpha,\sigma}^2\right).$$

Note that $|Z|^2 = \Re(Z)^2 + \Im(Z)^2 \sim \chi^2_2$. It follows from Lemma 1 of [50] that

$$N_\sigma\mathbb{P}\left(|Z|^2 > 1 + 2\sqrt{\log\frac{N_\sigma}{\alpha}} + 2\log\frac{N_\sigma}{\alpha}\right) \leq N_\sigma\exp\left(-\log\frac{N_\sigma}{\alpha}\right) = \alpha.$$

Thus, $\phi_{\alpha,\sigma}$ is indeed a level α test. As above, we must now show that the maximal type II error of $\phi_{\alpha,\sigma}$ vanishes. For some random variables $Z_{\sigma,k} \overset{\text{i.i.d.}}{\sim} \mathcal{CN}(0,2)$, $k \in I_\sigma$ and $Z_\sigma \sim \mathcal{CN}(0,2)$, we find

$$\begin{aligned}
\sup_{k\in I_\sigma}\sup_{|\delta|\geq\mu_\sigma} \mathbb{P}_{\delta f_k}\left(\phi_{\alpha,\sigma}(y_\sigma)=0\right) &= \sup_{k\in I_\sigma}\sup_{|\delta|\geq\mu_\sigma} \mathbb{P}_{\delta f_k}\left(\sup_{k'\in I_\sigma}\frac{|\langle Y_\sigma, Au_{k'}\rangle_{\mathcal{Y}}|}{\sigma\|Au_{k'}\|_{\mathcal{Y}}} \leq c_{\alpha,\sigma}\right) \\
&\leq \sup_{k\in I_\sigma}\sup_{|\delta|\geq\mu_\sigma} \mathbb{P}\left(\sup_{k'\in I_\sigma}\left|Z_{\sigma,k'} + \frac{\delta\langle Au_{k'}, Au_k\rangle_{\mathcal{Y}}}{\sigma\|Au_{k'}\|_{\mathcal{Y}}\|Au_k\|_{\mathcal{Y}}}\right|^2 \leq c_{\alpha,\sigma}^2\right) \\
&\leq \sup_{k\in I_\sigma}\sup_{|\delta|\geq\mu_\sigma} \mathbb{P}\left(\left|Z_{\sigma,k} + \frac{\delta}{\sigma}\right|^2 \leq c_{\alpha,\sigma}^2\right) = \sup_{|\delta|\geq\mu_\sigma} \mathbb{P}\left(\left|Z_\sigma + \frac{\delta}{\sigma}\right|^2 \leq c_{\alpha,\sigma}^2\right)
\end{aligned}$$

We have

$$\left|Z_\sigma + \frac{\delta}{\sigma}\right|^2 = |Z_\sigma|^2 + 2\Re\left(\frac{\bar{\delta}}{\sigma}Z_\sigma\right) + \frac{|\delta|^2}{\sigma^2} \geq \frac{2|\delta|}{\sigma}Z'_\sigma + \frac{|\delta|^2}{\sigma^2},$$

for some $Z'_\sigma \sim \mathcal{N}(0,1)$. It follows that

$$\sup_{|\delta|\geq\mu_\sigma} \mathbb{P}\left(\left|Z_\sigma + \frac{\delta}{\sigma}\right|^2 \leq c^2_{\alpha,\sigma}\right) \leq \mathbb{P}\left(\frac{2\mu_\sigma}{\sigma} Z'_\sigma + \frac{\mu_\sigma^2}{\sigma^2} \leq 1 + 2\sqrt{\log\frac{N_\sigma}{\alpha}} + 2\log\frac{N_\sigma}{\alpha}\right)$$
$$\leq \mathbb{P}\left(Z'_\sigma \leq \sqrt{\frac{1}{2}\log N_\sigma} - (1+\varepsilon_\sigma)\sqrt{\frac{1}{2}\log N_\sigma} + O(1)\right) \to 0,$$

since $\varepsilon_\sigma\sqrt{\log N_\sigma} \to \infty$. □

Proof of the lower bound

We suppose that $\mathcal{X}$ and $\mathcal{Y}$ are complex Hilbert spaces. The proof for the real case is analogous. In fact, the proof of Theorem 3.2 can in principle be derived from Proposition 4.10 and Lemma 7.2 from [40] with just a few adjustments.

Proof of Theorem 3.2. Let I^*_σ be the largest subset of I_σ such that $\Re(\langle Au_k, Au_{k'}\rangle_\mathcal{Y}) \leq 0$ for any distinct $k, k' \in I^*_\sigma$. Recall that $|I^*_\sigma| = N^*_\sigma$. Recall that

$$y_{\sigma,i} := \langle Y_\sigma, e_i\rangle_\mathcal{Y} = \langle Af, e_i\rangle_\mathcal{Y} + \sigma\xi_i,$$

where $\xi_i \overset{\text{i.i.d.}}{\sim} \mathcal{CN}(0,2)$ for $i \in \mathbb{N}$. This means that under H_0 the random sequence

$$\tilde{y}_\sigma := \left(\frac{1}{\sigma}\Re(y_{\sigma,1}), \frac{1}{\sigma}\Im(y_{\sigma,1}), \frac{1}{\sigma}\Re(y_{\sigma,2}), \frac{1}{\sigma}\Im(y_{\sigma,2}), \dots\right)$$

is a sequence of i.i.d. $\mathcal{N}(0,1)$-distributed random variables. Any test statistic may be expressed in terms of the Gaussian sequence $\tilde{y}_\sigma$. Hence, any test may be expressed as a function of $\tilde{y}_\sigma$.

Bayesian alternative. For $k \in I_\sigma$ we define $f_k := \frac{\mu_\sigma u_k}{\|Au_k\|_\mathcal{Y}}$, and let π_σ be the prior distribution on the alternative set $\mathcal{F}_\sigma(\mu_\sigma)$ given by $\pi_\sigma = \frac{1}{N^*_\sigma}\sum_{k\in I^*_\sigma}\delta_{f_k}$. The idea is to bound the maximal type II error probability from below by the mean (in terms of π_σ) type II error probability as follows:

$$\gamma_\sigma = \inf_{\phi\in\Phi_\sigma}\left[\alpha_\sigma(\phi) + \sup_{f\in\mathcal{F}_\sigma(\mu_\sigma)}\beta_\sigma(\phi,f)\right] \geq \inf_{\phi\in\Phi_\sigma}\left[\alpha_\sigma(\phi) + \frac{1}{N^*_\sigma}\sum_{k\in I^*_\sigma}\beta_\sigma(\phi,f_k)\right].$$

We may say that it suffices to analyze the "simpler" testing problem

$$H_0 : f \equiv 0 \quad \text{vs} \quad H_{\pi,\sigma} : f \sim \pi_\sigma,$$

in terms of its mean type II error, instead of (1.2) in the minimax sense. We have (cf. [40], chapter 2)

$$\inf_{\phi\in\Phi_\sigma}\left[\alpha_\sigma(\phi) + \frac{1}{N^*_\sigma}\sum_{k\in I^*_\sigma}\beta_\sigma(\phi,f_k)\right] = 1 - \frac{1}{2}\mathbb{E}_0\left|\mathbb{E}_{\pi_\sigma}\frac{d\mathbb{P}_f}{d\mathbb{P}_0}(\tilde{y}_\sigma) - 1\right|,$$

where $\mathbb{E}_{\pi_\sigma}\frac{d\mathbb{P}_f}{d\mathbb{P}_0}(\tilde{y}_\sigma) = \frac{1}{N^*_\sigma}\sum_{k\in I^*_\sigma}\frac{d\mathbb{P}_{f_k}}{d\mathbb{P}_0}(\tilde{y}_\sigma)$. We see that, in order to show indistinguishability, it suffices to show that

$$\mathbb{E}_0\left|\mathbb{E}_{\pi_\sigma}\frac{d\mathbb{P}_f}{d\mathbb{P}_0}(\tilde{y}_\sigma) - 1\right| \to 0. \tag{6.1}$$

We denote the distribution of $\tilde{y}_\sigma$ on $\mathbb{R}^\mathbb{N}$ under H_0 by $\mathbb{P}_0$ (this is of course the standard Gaussian distribution on $\mathbb{R}^\mathbb{N}$). The Cameron-Martin space corresponding to $(\mathbb{R}^\mathbb{N}, \mathbb{P}_0)$ is $\mathcal{H} = \ell^2$ with norm

$\|\cdot\|_{\mathcal{H}}^2 = \|\cdot\|_2^2$ (see for example Example 4.1 of [51]). If $f = f_k$, then $\tilde{y}_\sigma$ has distribution $\mathbb{P}_{f_k}$ defined by $\mathbb{P}_{f_k}(\cdot) = \mathbb{P}_0(\cdot - h_k)$, where $h_k \in \mathbb{R}^{\mathbb{N}}$ is given by

$$h_{k,2j-1} = \frac{\mu_\sigma}{\sigma\|Au_k\|_{\mathcal{Y}}}\Re(\langle Au_k, e_j\rangle_{\mathcal{Y}}), \quad h_{k,2j} = \frac{\mu_\sigma}{\sigma\|Au_k\|_{\mathcal{Y}}}\Im(\langle Au_k, e_j\rangle_{\mathcal{Y}}), \quad j \in \mathbb{N}.$$

It follows that

$$\|h_k\|_{\mathcal{H}}^2 = \sum_{i\in\mathbb{N}} \frac{\mu_\sigma^2}{\sigma^2\|Au_k\|_{\mathcal{Y}}^2}|\langle Au_k, e_i\rangle_{\mathcal{Y}}|^2 = \frac{\mu_\sigma^2}{\sigma^2},$$

which also shows that $h \in \mathcal{H}$. Thus, by the Cameron-Martin theorem (see Theorem 5.1 of [51]),

$$\begin{aligned}\frac{d\mathbb{P}_{f_k}}{d\mathbb{P}_0}(\tilde{y}_\sigma) &= \exp\left(\sum_{i\in\mathbb{N}} h_{k,i}\tilde{y}_{\sigma,i} - \frac{\mu_\sigma^2}{2\sigma^2}\right) \\ &= \exp\left(\frac{\mu_\sigma}{\sigma\|Au_k\|_{\mathcal{Y}}}\sum_{j\in\mathbb{N}}(\Re(\langle Au_k, e_j\rangle_{\mathcal{Y}})\tilde{y}_{\sigma,2j-1} + \Im(\langle Au_k, e_j\rangle_{\mathcal{Y}})\tilde{y}_{\sigma,2j}) - \frac{\mu_\sigma^2}{2\sigma^2}\right) \overset{H_0}{=} \exp\left(X_{\sigma,k}\right),\end{aligned}$$

where $X_{\sigma,k} \sim \mathcal{N}\left(-\frac{\mu_\sigma^2}{2\sigma^2}, \frac{\mu_\sigma^2}{\sigma^2}\right)$. Note that the distribution of $X_{\sigma,k}$ does not depend on k. However, the collection $\{X_{\sigma,k} : k \in I_\sigma\}$ is, in general, not independent. We have

$$\mathbb{E}_{\pi_\sigma}\frac{d\mathbb{P}_f}{d\mathbb{P}_0}(Y_\sigma) = \frac{1}{N_\sigma^*}\sum_{k\in I_\sigma^*}\exp(X_{\sigma,k}).$$

In order to show that (6.1) holds, we employ a weak law of large numbers, namely Theorem 3.2 from [71]. However, note that similar ideas have been used in [39].

A weak law of large numbers. Let $(\sigma_m)_{m\in\mathbb{N}}$ be a sequence of positive real numbers, such that $\sigma_m \searrow 0$ as $m \to \infty$. Consider the triangular array of random variables $\{\exp(X_{\sigma_m,k}) : m \in \mathbb{N}, k \in I_{\sigma_m}^*\}$. Note that

$$\mathrm{Cov}\left(\exp(X_{\sigma_m,k}), \exp(X_{\sigma_m,k'})\right) = \exp\left(\frac{\mu_{\sigma_m}^2\Re(\langle Au_k, Au_{k'}\rangle_{\mathcal{Y}})}{\sigma_m^2\|Au_k\|_{\mathcal{Y}}\|Au_{k'}\|_{\mathcal{Y}}}\right) - 1 \leq 0,$$

for any m and any two distinct $k, k' \in I_{\sigma_m}^*$. Let $(\kappa_m)_{m\in\mathbb{N}}$ be another sequence of real numbers given by $\kappa_m = (N_{\sigma_m}^*)^{(1+\varepsilon_{\sigma_m})(1-\varepsilon_{\sigma_m})^2}$. Then $\kappa_m \to \infty$ as $m \to \infty$ and

$$(N_{\sigma_m}^*)^{-1}\kappa_m = (N_{\sigma_m}^*)^{-\varepsilon_{\sigma_m}+O(\varepsilon_{\sigma_m}^2)} \to 0.$$

It follows that

$$\begin{aligned}\frac{1}{N_{\sigma_m}^*}\sum_{k\in I_{\sigma_m}^*}\mathbb{E}_0\left[(\exp(X_{\sigma_m,k}))\,\mathbb{1}(\exp(X_{\sigma_m,k}) > \kappa_m)\right] &= \mathbb{P}\left[\mathcal{N}(0,1) \leq \frac{\mu_{\sigma_m}}{2\sigma_m} - \frac{\sigma_m\log\kappa_m}{\mu_{\sigma_m}}\right] \\ &\leq \mathbb{P}\left[\mathcal{N}(0,1) \leq -\varepsilon_{\sigma_m}(1+\varepsilon_{\sigma_m})\sqrt{\frac{1}{2}\log N_{\sigma_m}^*}\right],\end{aligned}$$

which vanishes as $m \to \infty$, since $\varepsilon_\sigma\sqrt{2\log N_\sigma^*} \to \infty$. We can now employ Theorem 3.2 from [71], which immediately yields that

$$\mathbb{E}_0\left|\mathbb{E}_{\pi_\sigma}\frac{d\mathbb{P}_f}{d\mathbb{P}_0}(Y_\sigma) - 1\right| = \mathbb{E}_0\left|\frac{1}{N_{\sigma_m}^*}\sum_{k\in I_{\sigma m}^*}\exp(X_{\sigma_m,k}) - 1\right| \to 0,$$

as $m \to \infty$.

□

Remaining proofs

Proof of Corollary 3.3. For $k \in I_\sigma$, let

$$S_\sigma(k) = \{k' \in I_\sigma : \Re(\langle Au_k, Au_{k'}\rangle_{\mathcal{Y}}) \geq 0\}.$$

We construct a subset I'_σ of I_σ iteratively as follows. We choose $k_1 \in I_\sigma$ arbitrarily, then choose $k_2 \in I_\sigma \setminus S_\sigma(k_1)$ arbitrarily, then choose $k_3 \in I_\sigma \setminus (S_\sigma(k_1) \cup S_\sigma(k_2))$ arbitrarily, and continue until $S_\sigma(k_1) \cup S_\sigma(k_2) \cup \ldots = I_\sigma$. Then set $I'_\sigma = \{k_1, k_2, k_3, \ldots\}$. Since, by assumption, $|S_\sigma(k)| \leq M_\sigma$ for any $k \in I_\sigma$, it follows that $|I'_\sigma| \geq N_\sigma/M_\sigma \gtrsim N_\sigma^{1+\varepsilon_\sigma}$. Since the set I^*_σ can be constructed as above, Theorem 3.2 yields

$$\mu^*_\sigma \gtrsim (1-\varepsilon_\sigma)\sqrt{1+\varepsilon_\sigma}\sqrt{2\log N_\sigma}.$$

Thus,

$$\sqrt{1-\varepsilon_\sigma-\varepsilon_\sigma^2+\varepsilon_\sigma^3} \precsim \frac{\mu^*_\sigma}{\sqrt{2\log N_\sigma}} \precsim 1+\varepsilon_\sigma,$$

and the claim follows. □

Proof of Corollary 3.4. Since $A\zeta_k = s_k\eta_k$ for all k, and the system $(\eta_k)_{k\in\mathbb{N}}$ is orthonormal, this follows immediately from Corollary 3.3. □

Proof of Lemma 3.7. Since, by assumption, $\mu_\sigma/\sigma \to \infty$, we can choose a family of positive integers $(n_\sigma)_{\sigma>0}$, such that $n_\sigma \to \infty$ as $\sigma \to 0$ and

$$\frac{\mu_\sigma}{\sigma} - \sqrt{2\log n_\sigma} \to \infty.$$

For $m \in \{0, \ldots, n_\sigma - 1\}$ let

$$w_{\sigma,m} = 2^{j_\sigma/2}\psi\left(2^{j_\sigma}(\cdot - m/n_\sigma)\right),$$

and let $w^{(per)}_{\sigma,m} = \sum_{z\in\mathbb{Z}} w_{\sigma,m}(\cdot + z)$. For $\alpha \in (0,1)$ consider the test $\phi_{\alpha,\sigma}(y_\sigma) = \mathbb{1}\{T_\sigma > c_{\alpha,\sigma}\}$ with threshold $c_{\alpha,\sigma} = \sqrt{2\log(n_\sigma/\alpha)}$ and test statistic

$$T_\sigma = \sup_{0\leq m\leq n_\sigma-1} \frac{\left|\langle Y_\sigma, Aw^{(per)}_{\sigma,m}\rangle_{\mathcal{Y}}\right|}{\sigma\|Aw^{(per)}_{\sigma,m}\|_{\mathcal{Y}}}.$$

It is easy to see that $\phi_{\alpha,\sigma}$ is a level α test. Let $f_l = \frac{\psi^{(per)}_{j_\sigma,l}}{\|A\psi^{(per)}_{j_\sigma,l}\|_{\mathcal{Y}}}$. For $l \in \{0, \ldots, 2^{j_\sigma} - 1\}$ we define $m^*(l) = \arg\min\{|2^{-j_\sigma}l - n_\sigma^{-1}m| : m \in \{0, \ldots, n_\sigma - 1\}\}$. As in the previous proofs we find

$$\sup_{0\leq l\leq 2^{j_\sigma}-1} \sup_{\delta\geq\mu_\sigma} \mathbb{P}_{\delta f_l}\left(\phi_{\alpha,\sigma}(y_\sigma) = 0\right) \leq \mathbb{P}\left(Z \leq c_{\alpha,\sigma} - \frac{\mu_\sigma}{\sigma} \inf_{0\leq l\leq 2^{j_\sigma}-1} \frac{\left\langle A\psi^{(per)}_{j_\sigma,l}, Aw^{(per)}_{j_\sigma,m^*(l)}\right\rangle_{\mathcal{Y}}}{\|A\psi^{(per)}_{j_\sigma,l}\|_{\mathcal{Y}}\|Aw^{(per)}_{\sigma,m^*(l)}\|_{\mathcal{Y}}}\right),$$

for some $Z \sim \mathcal{N}(0,1)$. It remains to show that

$$\inf_{0\leq l\leq 2^{j_\sigma}-1} \frac{\left\langle A\psi^{(per)}_{j_\sigma,l}, Aw^{(per)}_{j_\sigma,m^*(l)}\right\rangle_{\mathcal{Y}}}{\|A\psi^{(per)}_{j_\sigma,l}\|_{\mathcal{Y}}\|Aw^{(per)}_{\sigma,m^*(l)}\|_{\mathcal{Y}}} \to 1,$$

as $\sigma \to 0$. Note that, due to periodicity, for any $l \in \{0, \ldots, 2^{j_\sigma} - 1\}$ and $m \in \{0, \ldots, n_\sigma - 1\}$,

$$\begin{aligned}\|A\psi^{(per)}_{j_\sigma,l}\|^2_{\mathcal{Y}} = \|Aw^{(per)}_{\sigma,m}\|^2_{\mathcal{Y}} = \|A\psi^{(per)}_{j_\sigma,0}\|^2_{\mathcal{Y}} &= \int_0^1 \left(\int_0^1 h(u-x)\psi^{(per)}_{j_\sigma,0}(u)du\right)^2 dx \\ &= \int_0^1 \left(\int_{-\infty}^{\infty} h(u-x)2^{j_\sigma/2}\psi(2^{j_\sigma}u)du\right)^2 dx.\end{aligned}$$

It follows from equation (6.15) of [52], that for any $x \in \mathbb{R}$,

$$\lim_{\sigma \to 0} 2^{3j_\sigma/2} \int_{-\infty}^{\infty} h(u-x) 2^{j_\sigma/2} \psi\left(2^{j_\sigma} u\right) du = Ch'(x),$$

for some constant $C > 0$. Due to Theorem 6.2 of [52], there exist an integrable function θ, such that $-\frac{d}{dx}\theta(x) = \psi(x)$. Since ψ has bounded support, θ has bounded support as well. Thus, we find by substituting and integrating by parts that

$$\begin{aligned} 2^{3j_\sigma/2} \int_{-\infty}^{\infty} h(u-x) 2^{j_\sigma/2} \psi\left(2^{j_\sigma} u\right) du &= 2^{3j_\sigma/2} \int_{-\infty}^{\infty} h(2^{-j_\sigma} v - x) 2^{-j_\sigma/2} \psi\left(v\right) dv \\ &= -\int_{-\infty}^{\infty} h'(2^{-j_\sigma} v - x) \theta\left(v\right) dv. \end{aligned}$$

Since, by assumption, h' is Lipschitz and periodic, it follows that it is bounded. Thus, for any $x \in \mathbb{R}$,

$$\left| 2^{3j_\sigma/2} \int_{-\infty}^{\infty} h(u-x) 2^{j_\sigma/2} \psi\left(2^{j_\sigma} u\right) du \right| \leq C' \int_{-\infty}^{\infty} |\theta\left(v\right)| dv.$$

It follows from the dominated convergence theorem that

$$2^{3j_\sigma/2} \| A\psi_{j_\sigma,0}^{(per)} \|_{\mathcal{Y}} \to C \| h' \|_{\mathcal{Y}},$$

as $\sigma \to \infty$. We have

$$\frac{\left\langle A\psi_{j_\sigma,l}^{(per)}, Aw_{j_\sigma,m^*(l)}^{(per)} \right\rangle_{\mathcal{Y}}}{\| A\psi_{j_\sigma,l}^{(per)} \|_{\mathcal{Y}} \| Aw_{\sigma,m^*(l)}^{(per)} \|_{\mathcal{Y}}} = 1 - \frac{\| A\psi_{j_\sigma,l}^{(per)} - Aw_{j_\sigma,m^*(l)}^{(per)} \|_{\mathcal{Y}}^2}{\| A\psi_{j_\sigma,0}^{(per)} \|_{\mathcal{Y}}^2}.$$

We substitute twice, integrate by parts, use that h' is Lipschitz and that $\left|2^{-j_\sigma} l - m^*(l) n_\sigma^{-1}\right| \leq \frac{1}{2n_\sigma}$, for any $k \in I_\sigma$, which follows immediately from the definition of $m^*(l)$, to obtain

$$\begin{aligned} &\left| A\psi_{j_\sigma,l}^{(per)}(x) - Aw_{j_\sigma,m^*(l)}^{(per)}(x) \right| \\ &= \left| \int_{-\infty}^{\infty} h(u-x) 2^{j_\sigma/2} \psi(2^{j_\sigma}(u - 2^{-j_\sigma} l)) du - \int_{-\infty}^{\infty} h(u-x) 2^{j_\sigma/2} \psi(2^{j_\sigma}(u - m^*(l) n_\sigma^{-1})) du \right| \\ &= \left| \int_{-\infty}^{\infty} \left[h(2^{-j_\sigma} v + 2^{-j_\sigma} l - x) - h(2^{-j_\sigma} v + m^*(l) n_\sigma^{-1} - x) \right] 2^{-j_\sigma/2} \psi(v) dv \right| \\ &= 2^{-3j_\sigma/2} \left| \int_{-\infty}^{\infty} \left[h'(2^{-j_\sigma} v + 2^{-j_\sigma} l - x) - h'(2^{-j_\sigma} v + m^*(l) n_\sigma^{-1} - x) \right] \theta(v) dv \right| \\ &\leq 2^{-3j_\sigma/2} C'' \left| 2^{-j_\sigma} l - m^*(l) n_\sigma^{-1} \right| \int_{-\infty}^{\infty} |\theta(v)| dv \leq C''' \frac{2^{-3j_\sigma/2}}{2n_\sigma}. \end{aligned}$$

It follows that

$$\inf_{0 \leq l \leq 2^{j_\sigma} - 1} \frac{\| A\psi_{j_\sigma,l}^{(per)} - Aw_{j_\sigma,m^*(l)}^{(per)} \|_{\mathcal{Y}}^2}{\| A\psi_{j_\sigma,0}^{(per)} \|_{\mathcal{Y}}^2} \leq C'''' \frac{n_\sigma^{-2}}{2^{3j_\sigma} \| A\psi_{j_\sigma,0}^{(per)} \|_{\mathcal{Y}}^2} \to 0,$$

since $n_\sigma \to \infty$ and the denominator converges to a positive constant.

□

6.2 Proofs for section 3.2

Techniques used in the following proofs are inspired by [49]. Note that here we only consider the case that $\mathcal{X}$ and $\mathcal{Y}$ are complex spaces. The proofs for the case that they are real is analogous.

Proof of the nonasymptotic upper bound

Proof of Theorem 3.9. Define the test

$$\phi_{\alpha,\sigma}(y_\sigma) = \mathbb{1}\left\{T_\sigma > t_{1-\alpha,\sigma}\right\}, \tag{6.2}$$

where $T_\sigma := T_\sigma(Y_\sigma) := \sum_{k\in I_\sigma} |\langle Y_\sigma, v_k\rangle_{\mathcal{Y}}|^2$, and $t_{1-\alpha,\sigma}$ is the $(1-\alpha)$-quantile of T_σ (which follows a generalized χ^2-distribution) under H_0. Thus, by its very definition, $\phi_{\alpha,\sigma}$ is a level α test. We need to show that, if ν_σ is large enough, for any $f \in \mathcal{F}_\sigma^L(\nu_\sigma)$

$$\mathbb{P}_f\left(T_\sigma \le t_{1-\alpha,\sigma}\right) \le \delta - \alpha. \tag{6.3}$$

We aim to show that asymptotically $t_{1-\alpha,\sigma} \le t_{\delta-\alpha,\sigma}(f)$ whenever $\nu_\sigma \ge \sqrt{2}d_\alpha(\delta)\sigma\sqrt{\|\Xi_{I_\sigma}\|_F}$, where $t_{\delta-\alpha,\sigma}(f)$ denotes the $\delta-\alpha$ quantile of T_σ when f is the true underlying signal. First, we need to discuss the distribution of T_σ.
For $f \in \mathcal{F}_\sigma^L(\nu_\sigma)$, the random vector $(\langle Y_\sigma, v_k\rangle_{\mathcal{Y}})_{k\in I_\sigma}$ is normally distributed with with mean vector $m_\sigma = (\lambda_k\langle f, u_k\rangle_{\mathcal{X}})_{k\in I_\sigma}$ and covariance matrix $2\sigma^2\Xi_{I_\sigma}$. Since Ξ_σ is Hermitian and positive definite by assumption, it can be decomposed as

$$\Xi_{I_\sigma} = U_\sigma S_\sigma U_\sigma^H,$$

where U is unitary and S_σ is a diagonal matrix containing the (real and positive) eigenvalues $(s_k)_{k\in I_\sigma}$ of Ξ_σ. It follows that the random vector $(\langle Y_\sigma, v_k\rangle_{\mathcal{Y}})_{k\in I_\sigma}$ can be written as $\sqrt{2}\sigma U_\sigma S_\sigma^{1/2} Z_\sigma + m_\sigma$ for some $Z_\sigma \sim \mathcal{CN}(0, \mathrm{id}_{N_\sigma})$ and thus,

$$\begin{aligned}
T_\sigma &= (\sqrt{2}\sigma U_\sigma S_\sigma^{1/2} Z_\sigma + m_\sigma)^H (\sqrt{2}\sigma U_\sigma S_\sigma^{1/2} Z_\sigma + m_\sigma) \\
&= 2\sigma^2 (Z_\sigma + (\sqrt{2}\sigma)^{-1} S_\sigma^{-1/2} U_\sigma^H m_\sigma)^H S_\sigma (Z_\sigma + (\sqrt{2}\sigma)^{-1} S_\sigma^{-1/2} U_\sigma^H m_\sigma) \\
&= 2\sigma^2 \sum_{k\in I_\sigma} s_k \left| Z_{\sigma,k} - \frac{1}{\sqrt{2}}\tilde{m}_{\sigma,k} \right|^2 \\
&= 2\sigma^2 \sum_{k\in I_\sigma} s_k \left[\left(\Re(Z_{\sigma,k}) - \frac{1}{\sqrt{2}}\Re(\tilde{m}_{\sigma,k}) \right)^2 + \left(\Im(Z_{\sigma,k}) - \frac{1}{\sqrt{2}}\Im(\tilde{m}_{\sigma,k}) \right)^2 \right] \\
&= \sigma^2 \sum_{k\in I_\sigma} s_k \left[(Z'_k - \Re(\tilde{m}_{\sigma,k}))^2 + (Z''_k - \Im(\tilde{m}_{\sigma,k}))^2 \right],
\end{aligned}$$

where $\tilde{m}_\sigma = \sigma^{-1} S_\sigma^{-1/2} U_\sigma^H m_\sigma$ and $Z', Z'' \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \mathrm{id}_{N_\sigma})$. In other words, T_σ is the sum of $2N_\sigma$ weighted non-central chi-squared random variables. Note that

$$\sigma^2 \sum_{k\in I_\sigma} s_k |\tilde{m}_{\sigma,k}|^2 = (S_\sigma^{-1/2} U_\sigma^H m_\sigma)^H S (S_\sigma^{-1/2} U_\sigma^H m_\sigma) = m_\sigma^H m_\sigma = \sum_{k\in I_\sigma} |\lambda_k \langle f, u_k\rangle_{\mathcal{X}}|^2.$$

Upper bound for $t_{1-\alpha,\sigma}$. Under H_0 we have that $T_\sigma = \sigma^2 \sum_{k\in I_\sigma} s_k \left[(Z'_k)^2 + (Z''_k)^2 \right]$. It follows from Lemma 1 from [50] that for any $t > 0$

$$\mathbb{P}_0\left(T_\sigma > 2\sigma^2 \sum_{k\in I_\sigma} s_k + 2\sigma^2 \sqrt{2t \sum_{k\in I_\sigma} s_k^2} + \sigma^2 t \left(\sup_{k\in I_\sigma} s_k \right) \right) \le \exp(-t),$$

and thus, setting $t = \log(1/\alpha)$, we have

$$t_{1-\alpha,\sigma} \le 2\sigma^2 \sum_{k\in I_\sigma} s_k + 2\sqrt{2\log(1/\alpha)}\sigma^2 \|\Xi_{I_\sigma}\|_F + \log(1/\alpha)\sigma^2 \sup_{k\in I_\sigma} s_k.$$

Lower bound for $t_{\delta-\alpha,\sigma}(f)$. We use Lemma 2 from [49], which yields that for any $t > 0$

$$\mathbb{P}_f\left(T_\sigma \leq \mathbb{E}T_\sigma - 2\sqrt{2t}\sqrt{\sigma^4 \sum_{k\in I_\sigma} s_k^2 + \sigma^4 \sum_{k\in I_\sigma} s_k^2 |\tilde{m}_{\sigma,k}|^2}\right) \leq \exp(-t).$$

Setting $t = \log(1/(\delta-\alpha))$, it follows that

$$t_{\delta-\alpha,\sigma}(f) \geq \mathbb{E}T_\sigma - 2\sqrt{2\log(1/(\delta-\alpha))}\sqrt{\sigma^4\|\Xi_{I_\sigma}\|_F^2 + \sigma^4 \sum_{k\in I_\sigma} s_k^2|\tilde{m}_{\sigma,k}|^2}$$

$$\geq 2\sigma^2 \sum_{k\in I_\sigma} s_k + \sum_{k\in I_\sigma} |\lambda_k\langle f, u_k\rangle_{\mathcal{X}}|^2 - 2\sqrt{2\log(1/(\delta-\alpha))}\left[\sigma^2\|\Xi_{I_\sigma}\|_F + \sigma\sqrt{\|\Xi_{I_\sigma}\|_F}\sqrt{\sum_{k\in I_\sigma}|\lambda_k\langle f, u_k\rangle_{\mathcal{X}}|^2}\right],$$

where we used that $\sup_{k\in I_\sigma} s_k \leq \sqrt{\sum_{k\in I_\sigma} s_k^2} = \|\Xi_{I_\sigma}\|_F$.

Comparing the bounds. It follows that $t_{1-\alpha,\sigma} \leq t_{\delta-\alpha,\sigma}(f)$ is true when

$$\sum_{k\in I_\sigma} |\lambda_k\langle f, u_k\rangle_{\mathcal{X}}|^2 - 2\sigma\sqrt{2\log(1/(\delta-\alpha))}\sqrt{\|\Xi_{I_\sigma}\|_F}\sqrt{\sum_{k\in I_\sigma}|\lambda_k\langle f, u_k\rangle_{\mathcal{X}}|^2}$$
$$- \left(2\sqrt{2\log(1/(\delta-\alpha))} + 2\sqrt{2\log(1/\alpha)}\right)\sigma^2\|\Xi_{I_\sigma}\|_F - \log(1/\alpha)\sigma^2 \sup_{k\in I_\sigma} s_k \geq 0,$$

which holds when

$$\sqrt{\sum_{k\in I_\sigma}|\lambda_k\langle f, u_k\rangle_{\mathcal{X}}|^2} \geq \sqrt{2}\left[\sqrt{\log\frac{1}{\delta-\alpha}} + \left(\log\frac{1}{\alpha(\delta-\alpha)} + \sqrt{2\log\frac{1}{\delta-\alpha}} + \sqrt{2\log\frac{1}{\alpha}}\right)^{1/2}\right]\sigma\sqrt{\|\Xi_{I_\sigma}\|_F}.$$

□

Proof of the non-asymptotic lower bound

Proof of Theorem 3.10. The matrix $\tilde{\Xi}_{I_\sigma}$ given by $(\tilde{\Xi}_{I_\sigma})_{k,k'} = \langle \tilde{v}_k, \tilde{v}_{k'}\rangle_{\mathcal{Y}}$ is Hermitian and positive definite, and thus, and we have the decompositions

$$\tilde{\Xi}_{I_\sigma} = \tilde{U}_\sigma \tilde{S}_\sigma \tilde{U}_\sigma^H, \quad \tilde{\Xi}_{I_\sigma}^{-1} = \tilde{U}_\sigma \tilde{S}_\sigma^{-1} \tilde{U}_\sigma^H,$$

where $\tilde{U}_\sigma$ is unitary and $\tilde{S}_\sigma$ is a diagonal matrix with real and positive entries $(\tilde{s}_k)_{k\in I_\sigma}$ on its diagonal. The proof of the lower bound has the same core idea as the proof of Theorem 3.2: We start by defining a prior distribution on the set $\mathcal{F}_\sigma^L(\nu_\sigma)$. Let $w = (w_k)_{k\in I_\sigma}$ be a vector with $w_k \in \{-1, 1\}$ for all k, and define

$$\tilde{w} = \frac{\nu_\sigma}{\|\tilde{\Xi}_{I_\sigma}^{-1}\|_F}\tilde{U}_\sigma \tilde{S}_\sigma^{-1} w,$$

and

$$f_w = \sum_{k\in I_\sigma} \overline{\tilde{w}_k}\lambda_k^{-1} u_k.$$

Note that

$$\sum_{k\in I_\sigma} |\lambda_k\langle f_w, u_k\rangle_{\mathcal{X}}|^2 = \nu_\sigma^2,$$

and thus, indeed $f_w \in \mathcal{F}_\sigma^L(\nu_\sigma)$. As in the proof of Theorem 3.2 we get the likelihood ratio

$$\frac{d\mathbb{P}_{f_w}}{d\mathbb{P}_0}(\tilde{y}_\sigma) = \exp\left(\frac{1}{\sigma}\sum_{j\in\mathbb{N}}[\Re(\langle Af_w, e_j\rangle_{\mathcal{Y}})\tilde{y}_{\sigma,2j-1} + \Im(\langle Af_w, e_j\rangle_{\mathcal{Y}})\tilde{y}_{\sigma,2j}] - \frac{\|Af_w\|_{\mathcal{Y}}^2}{2\sigma^2}\right).$$

Note that $Af_w = \sum_{k\in I_\sigma}\overline{\tilde{w}_k}\tilde{v}_k$, an thus,

$$\|Af_w\|_{\mathcal{Y}}^2 = \sum_{k,l\in I_\sigma}\overline{\tilde{w}_k}\tilde{w}_l\langle\tilde{v}_k,\tilde{v}_l\rangle_{\mathcal{Y}} = \frac{\nu_\sigma^2}{\|\tilde{\Xi}_\sigma^{-1}\|_F^2}w^T\tilde{S}_\sigma^{-1}\tilde{U}_\sigma^H\tilde{\Xi}_{I_\sigma}\tilde{U}_\sigma\tilde{S}_\sigma^{-1}w = \frac{\nu_\sigma^2}{\|\tilde{\Xi}_{I_\sigma}^{-1}\|_F^2}\sum_{k\in I_\sigma}\frac{1}{\tilde{s}_k}.$$

Let $\tilde{v}^{(j)}$ be the vector with entries $\tilde{v}_k^{(j)} = \langle\tilde{v}_k, e_j\rangle_{\mathcal{Y}}$ for $k \in I_\sigma$. Then

$$\langle Af_w, e_j\rangle_{\mathcal{Y}} = \sum_{k\in I_\sigma}\overline{\tilde{w}_k}\langle\tilde{v}_k, e_j\rangle_{\mathcal{Y}} = \frac{\nu_\sigma}{\|\tilde{\Xi}_{I_\sigma}^{-1}\|_F}w^T\tilde{S}_\sigma^{-1}\tilde{U}_\sigma^H\tilde{v}^{(j)} = \frac{\nu_\sigma}{\|\tilde{\Xi}_{I_\sigma}^{-1}\|_F}\sum_{k\in I_\sigma}w_k\left(\tilde{S}_\sigma^{-1}\tilde{U}_\sigma^H\tilde{v}^{(j)}\right)_k,$$

and it follows that

$$\sum_{j\in\mathbb{N}}[\Re(\langle Af_w, e_j\rangle_{\mathcal{Y}})\tilde{y}_{\sigma,2j-1} + \Im(\langle Af_w, e_j\rangle_{\mathcal{Y}})\tilde{y}_{\sigma,2j}] = \frac{\nu_\sigma}{\|\tilde{\Xi}_{I_\sigma}^{-1}\|_F}\sum_{k\in I_\sigma}w_kZ_{\sigma,k},$$

where $Z_\sigma = (Z_{\sigma,k})_{k\in I_\sigma}$, with $Z_{\sigma,k} = \sum_{j\in\mathbb{N}}\left[\Re\left(\tilde{S}_\sigma^{-1}\tilde{U}_\sigma^H\tilde{v}^{(j)}\right)_k\tilde{y}_{\sigma,2j-1} + \Im\left(\tilde{S}_\sigma^{-1}\tilde{U}_\sigma^H\tilde{v}^{(j)}\right)_k\tilde{y}_{\sigma,2j}\right]$, is a normally distributed random vector with mean 0 and covariance matrix Σ given by

$$\begin{aligned}
\Sigma_{k,l} &= \sum_{j\in\mathbb{N}}\Re\left[\overline{\left(\tilde{S}_\sigma^{-1}\tilde{U}_\sigma^H\tilde{v}^{(j)}\right)_k}\left(\tilde{S}_\sigma^{-1}\tilde{U}_\sigma^H\tilde{v}^{(j)}\right)_l\right] \\
&= \sum_{j\in\mathbb{N}}\Re\left[\overline{\left(\sum_{m\in I_\sigma}\left(\tilde{S}_\sigma^{-1}\tilde{U}_\sigma^H\right)_{k,m}\langle\tilde{v}_m, e_j\rangle_{\mathcal{Y}}\right)}\left(\sum_{m'\in I_\sigma}\left(\tilde{S}_\sigma^{-1}\tilde{U}_\sigma^H\right)_{l,m'}\langle\tilde{v}_{m'}, e_j\rangle_{\mathcal{Y}}\right)\right] \\
&= \Re\left[\sum_{m,m'\in\mathbb{N}}\overline{\left(\tilde{S}_\sigma^{-1}\tilde{U}_\sigma^H\right)_{k,m}}\left(\tilde{S}_\sigma^{-1}\tilde{U}_\sigma^H\right)_{l,m'}\langle\tilde{v}_{m'}, \tilde{v}_m\rangle_{\mathcal{Y}}\right] = \Re\left(\tilde{S}_\sigma^{-1}\tilde{U}_\sigma^H\tilde{\Xi}_\sigma\tilde{U}_\sigma\tilde{S}_\sigma^{-1}\right)_{k,l} = \frac{1}{\tilde{s}_k}\delta_{k,l}.
\end{aligned}$$

In other words, the random variables $Z_{\sigma,k}$ are independent and $Z_{\sigma,k} \sim \mathcal{N}(0, 1/\tilde{s}_k)$ for $k \in I_\sigma$. Thus, under H_0, we have

$$\frac{d\mathbb{P}_{f_w}}{d\mathbb{P}_0}(\tilde{y}_\sigma) = \exp\left[-\frac{\nu_\sigma^2}{2\sigma^2\|\tilde{\Xi}_{I_\sigma}^{-1}\|_F^2}\sum_{k\in I_\sigma}\frac{1}{\tilde{s}_k}\right]\prod_{k\in I_\sigma}\exp\left[\frac{\nu_\sigma}{\sigma\|\tilde{\Xi}_{I_\sigma}^{-1}\|_F}\cdot\frac{w_k}{\sqrt{\tilde{s}_k}}Z'_{\sigma,k}\right],$$

where $Z'_{\sigma,k} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ for $k \in I_\sigma$.
Now, assume that $\hat{w}_k$, $k \in I_\sigma$ are independent Rademacher variables (which means that $\mathbb{P}(\hat{w}_k = 1) = \mathbb{P}(\hat{w}_k = -1) = 1/2$ for any k), that are also independent from $\tilde{y}_\sigma$, and let $\hat{w} = (\hat{w}_k)_{k\in I_\sigma}$ be the corresponding random vector. We denote by π_σ the (finitely supported) distribution of the random function $f_{\hat{w}}$ on $\mathcal{F}_\sigma^L(\nu_\sigma)$. As in the proof of Theorem 3.2 we have

$$\gamma_\sigma = \inf_{\phi\in\Phi_\sigma}\left[\alpha_\sigma(\phi) + \sup_{f\in\mathcal{F}_\sigma^L(\nu_\sigma)}\beta_\sigma(\phi, f)\right] = 1 - \frac{1}{2}\mathbb{E}_0\left|\mathbb{E}_{\pi_\sigma}\frac{d\mathbb{P}_f}{d\mathbb{P}_0}(Y_\sigma) - 1\right|,$$

Note that

$$\mathbb{E}_0\mathbb{E}_{\pi_\sigma}\frac{d\mathbb{P}_f}{d\mathbb{P}_0}(\tilde{y}_\sigma) = \mathbb{E}_{\pi_\sigma}\mathbb{E}_0\frac{d\mathbb{P}_f}{d\mathbb{P}_0}(\tilde{y}_\sigma) = 1,$$

and it follows that

$$\left(\mathbb{E}_0\left|\mathbb{E}_{\pi_\sigma}\frac{d\mathbb{P}_f}{d\mathbb{P}_0}(\tilde{y}_\sigma) - 1\right|\right)^2 \le \mathbb{E}_0\left(\mathbb{E}_{\pi_\sigma}\frac{d\mathbb{P}_f}{d\mathbb{P}_0}(\tilde{y}_\sigma)\right)^2 - 1,$$

and thus,

$$\gamma_\sigma(\nu_\sigma) \geq 1 - \frac{1}{2}\left(\mathbb{E}_0\left(\mathbb{E}_{\pi_\sigma}\frac{d\mathbb{P}_f}{d\mathbb{P}_0}(\tilde{y}_\sigma)\right)^2 - 1\right)^{1/2}.$$

We have

$$\mathbb{E}_{\pi_\sigma}\frac{d\mathbb{P}_f}{d\mathbb{P}_0}(\tilde{y}_\sigma) = \exp\left[-\frac{\nu_\sigma^2}{2\sigma^2\|\tilde{\Xi}_{I_\sigma}^{-1}\|_F^2}\sum_{k\in I_\sigma}\frac{1}{\tilde{s}_k}\right]\prod_{k\in I_\sigma}\cosh\left[\frac{\nu_\sigma}{\sigma\sqrt{\tilde{s}_k}\|\tilde{\Xi}_{I_\sigma}^{-1}\|_F}Z'_{\sigma,k}\right].$$

It follows that

$$\mathbb{E}_0\left(\mathbb{E}_{\pi_\sigma}\frac{d\mathbb{P}_f}{d\mathbb{P}_0}(\tilde{y}_\sigma)\right)^2 = \prod_{k\in I_\sigma}\cosh\left[\frac{\nu_\sigma^2}{\sigma^2\tilde{s}_k\|\tilde{\Xi}_{I_\sigma}^{-1}\|_F^2}\right] \leq \prod_{k\in I_\sigma}\exp\left[\frac{\nu_\sigma^4}{\sigma^4\tilde{s}_k\|\tilde{\Xi}_{I_\sigma}^{-1}\|_F^4}\right] = \exp\left[\frac{\nu_\sigma^4}{\sigma^4\|\tilde{\Xi}_{I_\sigma}^{-1}\|_F^2}\right],$$

where we used that $\mathbb{E}\cosh^2(tX) = \exp(t^2)\cosh(t^2)$ for any $t \in \mathbb{R}$ and $X \sim \mathcal{N}(0,1)$ and that $\cosh(t) \leq \exp(t^2/2)$ for any $t \in \mathbb{R}$. The claim follows immediately.

□

Remaining proofs

Proof of Corollary 3.11. The second part follows immediately from Theorem 3.10. For the first part, note that $\gamma_\sigma(\nu_\sigma) \leq \gamma_{\sigma,\alpha}(\nu_\sigma)$ for any α. Thus, the first part follows immediately from Theorem 3.9. □

Proof of Lemma 3.13. (1) Let $z = (z_k)_{k\in I_\sigma}$ be a non-zero complex vector. Then it follows from the fact that $(v_k)_{k\in I_\sigma}$ is a Riesz sequence that

$$z^H\Xi_{I_\sigma}z = \left\|\sum_{k\in I_\sigma}\overline{z_k}v_k\right\|_{\mathcal{Y}} \geq \left(C\sum_{k\in I_\sigma}|\overline{z_k}|^2\right)^{1/2} > 0,$$

for some constant $C > 0$. The proof for $\tilde{\Xi}_{I_\sigma}$ is analogous.
(2) The results of Theorem 3.9 and 3.10 (which can be applied since Ξ_{I_σ} and $\tilde{\Xi}_{I_\sigma}$ are positive definite) imply that $\|\tilde{\Xi}_\sigma^{-1}\|_F = O(\|\Xi_{I_\sigma}\|_F)$. It remains to show that $\|\Xi_{I_\sigma}\|_F \leq C\|\tilde{\Xi}_\sigma^{-1}\|_F$ for some constant $C > 0$. We have

$$\|\Xi_{I_\sigma}\|_F = \|\Xi_{I_\sigma}\tilde{\Xi}_{I_\sigma}\tilde{\Xi}_{I_\sigma}^{-1}\|_F \leq \|\Xi_{I_\sigma}\tilde{\Xi}_{I_\sigma}\|_2\|\tilde{\Xi}_{I_\sigma}^{-1}\|_F,$$

where $\|\Xi_{I_\sigma}\tilde{\Xi}_{I_\sigma}\|_2 = \max_{\|z\|_2=1}\|\Xi_{I_\sigma}\tilde{\Xi}_{I_\sigma}z\|_2$, where $\|\cdot\|_2$ denotes the euclidean norm on $\mathbb{C}^{N_\sigma}$. Now let $z = (z_k)_{k\in I_\sigma}$ be a complex vector with $\|z\|_2 = 1$. Recall that, since $(v_k)_{k\in I}$ and $(\tilde{v}_k)_{k\in I}$ are Riesz sequences, they are also frames of their respective spans. It follows that

$$\begin{aligned}
\|\Xi_{I_\sigma}\tilde{\Xi}_{I_\sigma}z\|_2^2 &= \sum_{k\in I_\sigma}\left|\sum_{l\in I_\sigma}\left(\sum_{j\in I_\sigma}\langle v_k, v_j\rangle_{\mathcal{Y}}\langle\tilde{v}_j,\tilde{v}_l\rangle_{\mathcal{Y}}\right)z_l\right|^2 \leq \sum_{k\in I}\left|\left\langle\sum_{j\in I_\sigma}\left\langle\sum_{l\in I_\sigma}\overline{z_l}\tilde{v}_l,\tilde{v}_j\right\rangle_{\mathcal{Y}}v_j, v_k\right\rangle_{\mathcal{Y}}\right|^2 \\
&\leq C\left\|\sum_{j\in I_\sigma}\left\langle\sum_{l\in I_\sigma}\overline{z_l}\tilde{v}_l,\tilde{v}_j\right\rangle_{\mathcal{Y}}v_j\right\|_{\mathcal{Y}}^2 \leq C'\sum_{j\in I_\sigma}\left|\left\langle\sum_{l\in I_\sigma}\overline{z_l}\tilde{v}_l,\tilde{v}_j\right\rangle_{\mathcal{Y}}\right|^2 \\
&\leq C'\sum_{j\in I}\left|\left\langle\sum_{l\in I_\sigma}\overline{z_l}\tilde{v}_l,\tilde{v}_j\right\rangle_{\mathcal{Y}}\right|^2 \leq C''\left\|\sum_{l\in I_\sigma}\overline{z_l}\tilde{v}_l\right\|_{\mathcal{Y}}^2 \leq C'''\sum_{l\in I_\sigma}|\overline{z_l}|^2 = C''',
\end{aligned}$$

which concludes this proof.
(3) Note that there are constants $C_1, C_2 > 0$, such that $C_1 \leq \|v_k\|_{\mathcal{Y}} \leq C_2$, for any $k \in I$, since $(v_k)_{k\in I}$ is a Riesz sequence. It follows that

$$\|\Xi_{I_\sigma}\|_F^2 = \sum_{k\in I_\sigma} \sum_{k'\in I_\sigma} |\langle v_k, v_{k'}\rangle_{\mathcal{Y}}|^2 \geq \sum_{k\in I_\sigma} |\langle v_k, v_k\rangle_{\mathcal{Y}}|^2 = \sum_{k\in I_\sigma} \|v_k\|_{\mathcal{Y}}^4 \geq C_1^4 N_\sigma,$$

and

$$\|\Xi_{I_\sigma}\|_F^2 \leq \sum_{k\in I_\sigma} \sum_{k'\in I} |\langle v_k, v_{k'}\rangle_{\mathcal{Y}}|^2 \leq C \sum_{k\in I_\sigma} \|v_k\|_{\mathcal{Y}}^2 \leq C C_2^2 N_\sigma.$$

The claim follows. □

Bibliography

[1] F. Abramovich and B. W. Silverman. Wavelet decomposition approaches to statistical inverse problems. *Biometrika*, 85(1):115–129, 1998.

[2] A. Antoniadis and J. Bigot. Poisson inverse problems. *Ann. Statist.*, 34(5):2132–2158, 2006.

[3] A. Aue and L. Horváth. Structural breaks in time series. *J. Time Series Anal.*, 34(1):1–16, 2013.

[4] F. Autin, M. Clausel, J.-M. Freyermuth, and C. Marteau. Maxiset point of view for signal detection in inverse problems. *Math. Methods Statist.*, 28(3):228–242, 2019.

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

[6] N. Bissantz, T. Hohage, A. Munk, and F. Ruymgaart. Convergence rates of general regularization methods for statistical inverse problems and applications. *SIAM J. Numer. Anal.*, 45(6):2610–2636, 2007.

[7] A. Böttcher and S. M. Grudsky. *Toeplitz matrices, asymptotic linear algebra, and functional analysis*. Birkhäuser Verlag, Basel, 2000.

[8] A. Böttcher and B. Silbermann. *Introduction to large truncated Toeplitz matrices*. Universitext. Springer-Verlag, New York, 1999.

[9] L. Boysen, A. Kempe, V. Liebscher, A. Munk, and O. Wittich. Consistencies and rates of convergence of jump-penalized least squares estimators. *Ann. Statist.*, 37(1):157–183, 2009.

[10] P. J. Brockwell and R. A. Davis. *Time series: theory and methods*. Springer Series in Statistics. Springer-Verlag, New York, second edition, 1991.

[11] B. E. Brodsky and B. S. Darkhovsky. *Nonparametric methods in change-point problems*, volume 243 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1993.

[12] E. Carlstein, H.-G. Müller, and D. Siegmund, editors. *Change-point problems*, volume 23 of *Lecture Notes–Monograph Series*. Institute of Mathematical Statistics, Hayward, CA, 1994.

[13] S. Chakar, E. Lebarbier, C. Lévy-Leduc, and S. Robin. A robust approach for estimating change-points in the mean of an AR(1) process. *Bernoulli*, 23(2):1408–1447, 2017.

[14] H. P. Chan and G. Walther. Detection with the scan and the average likelihood ratio. *Statist. Sinica*, 23(1):409–428, 2013.

[15] M. Csörgő and L. Horváth. *Limit theorems in change-point analysis*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester, 1997.

[16] I. Daubechies. *Ten lectures on wavelets*, volume 61 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992.

[17] H. Dette, T. Schüler, and M. Vetter. Multiscale change point detection for dependent data. *arXiv preprint arXiv:1811.05956*, 2018.

[18] D. L. Donoho. Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition. *Appl. Comput. Harmon. Anal.*, 2(2):101–126, 1995.

[19] L. Dümbgen and V. G. Spokoiny. Multiscale testing of qualitative hypotheses. *Ann. Statist.*, 29(1):124–152, 2001.

[20] L. Dümbgen and G. Walther. Multiscale inference about a density. *Ann. Statist.*, 36(4):1758–1785, 2008.

[21] A. Ebner, J. Frikel, D. Lorenz, J. Schwab, and M. Haltmeier. Regularization of Inverse Problems by Filtered Diagonal Frame Decomposition. *arXiv preprint, arXiv:2008.06219*, 2020.

[22] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems*, volume 375 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1996.

[23] F. Enikeeva, A. Munk, M. Pohlmann, and F. Werner. Bump detection in the presence of dependency: does it ease or does it load? *Bernoulli*, 26(4):3280–3310, 2020.

[24] F. Enikeeva, A. Munk, M. Pohlmann, and F. Werner. Supplement to "Dependent bump detection in the presence of dependency: Does it ease or does it load?". *DOI:10.3150/20-BEJ1226SUPP*, 2020.

[25] F. Enikeeva, A. Munk, and F. Werner. Bump detection in heterogeneous Gaussian regression. *Bernoulli*, 24(2):1266–1306, 2018.

[26] M. S. Ermakov. Minimax detection of a signal in Gaussian white noise. *Teor. Veroyatnost. i Primenen.*, 35(4):704–715, 1990.

[27] K. Frick, A. Munk, and H. Sieling. Multiscale change point inference. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 76(3):495–580, 2014.

[28] E. Giné and R. Nickl. *Mathematical foundations of infinite-dimensional statistical models*. Cambridge Series in Statistical and Probabilistic Mathematics, [40]. Cambridge University Press, New York, 2016.

[29] A. Goldenshluger, A. Juditsky, and A. Nemirovski. Hypothesis testing by convex optimization. *Electron. J. Stat.*, 9(2):1645–1712, 2015.

[30] A. Gut. The weak law of large numbers for arrays. *Statist. Probab. Lett.*, 14(1):49–52, 1992.

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

[32] M. Hanke. *A taste of inverse problems*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2017.

[33] T. Hotz, O. M. Schütte, H. Sieling, T. Polupanow, U. Diederichsen, C. Steinem, and A. Munk. Idealizing Ion Channel Recordings by a Jump Segmentation Multiresolution Filter. *IEEE Transactions on NanoBioscience*, 12(4):376–386, 2013.

[34] S. Hubmer and R. Ramlau. Frame decompositions of bounded linear operators in Hilbert spaces with applications in tomography. *Inverse Problems*, 37(5):Paper No. 055001, 30, 2021.

[35] I. A. Ibragimov and R. Z. Has'minskii. *Statistical estimation: asymptotic theory*. Springer Science & Business Media, 2013.

[36] I. A. Ibragimov and Yu. V. Linnik. *Independent and stationary sequences of random variables.* Wolters-Noordhoff Publishing, Groningen, 1971.

[37] Yu. Ingster, B. Laurent, and C. Marteau. Signal detection for inverse problems in a multidimensional framework. *Math. Methods Statist.*, 23(4):279–305, 2014.

[38] Yu. I. Ingster. Asymptotically minimax hypothesis testing for nonparametric alternatives. I-III. *Math. Methods Statist.*, 2(4):249–268, 1993.

[39] Yu. I. Ingster, T. Sapatinas, and I. A. Suslina. Minimax signal detection in ill-posed inverse problems. *Ann. Statist.*, 40(3):1524–1549, 2012.

[40] Yu. I. Ingster and I. A. Suslina. *Nonparametric goodness-of-fit testing under Gaussian models*, volume 169 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 2003.

[41] S. Jaffard. Propriétés des matrices "bien localisées" près de leur diagonale et quelques applications. *Ann. Inst. H. Poincaré Anal. Non Linéaire*, 7(5):461–476, 1990.

[42] X. J. Jeng, T. T. Cai, and H. Li. Optimal sparse segment identification with application in copy number variation analysis. *J. Amer. Statist. Assoc.*, 105(491):1156–1166, 2010.

[43] I. M. Johnstone, G. Kerkyacharian, D. Picard, and M. Raimondo. Wavelet deconvolution in a periodic setting. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 66(3):547–573, 2004.

[44] H. Keshavarz, C. Scott, and X. Nguyen. Optimal change point detection in Gaussian processes. *J. Statist. Plann. Inference*, 193:151–178, 2018.

[45] Hossein Keshavarz, Clayton Scott, and XuanLong Nguyen. Optimal change point detection in Gaussian processes. *arXiv preprint, arXiv:1506.01338*, 2017.

[46] J. Klosowski and J. Frahm. Image denoising for real-time MRI. *Magnetic resonance in medicine*, 77(3):1340–1352, 2017.

[47] T. Krivobokova, R. Briones, J. S. Hub, A. Munk, and B.L. de Groot. Partial Least-Squares Functional Mode analysis: Application to the Membrane Proteins AQP1, Aqy1, and CLC-ec1. *Biophysical Journal*, 103(4):786–796, 2012.

[48] B. Laurent, J.-M. Loubes, and C. Marteau. Testing inverse problems: a direct or an indirect problem? *J. Statist. Plann. Inference*, 141(5):1849–1861, 2011.

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

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

[51] M. Lifshits. *Lectures on Gaussian processes.* SpringerBriefs in Mathematics. Springer, Heidelberg, 2012.

[52] S. Mallat. *A wavelet tour of signal processing.* Academic Press, Inc., San Diego, CA, 1998.

[53] C. Marteau and P. Mathé. General regularization schemes for signal detection in inverse problems. *Math. Methods Statist.*, 23(3):176–200, 2014.

[54] A. Munk, T. Staudt, and F. Werner. Statistical foundations of nanoscale photonic imaging. *Nanoscale Photonic Imaging*, pages 125–143, 2020.

[55] F. Natterer. *The mathematics of computerized tomography.* B. G. Teubner, Stuttgart; John Wiley & Sons, Ltd., Chichester, 1986.

[56] E. Neher and B. Sakmann. *Single-channel recording.* Plenum Press, New York, 1995.

[57] M. Ordóñez Cabrera and A. I. Volodin. Mean convergence theorems and weak laws of large numbers for weighted sums of random variables under a condition of weighted integrability. *J. Math. Anal. Appl.*, 305(2):644–658, 2005.

[58] F. Pein, H. Sieling, and A. Munk. Heterogeneous change point inference. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 79(4):1207–1227, 2017.

[59] F. Pein, I. Tecuapetla-Gomez, O. M. Schütte, C. Steinem, and A. Munk. Fully-automatic multiresolution idealization for filtered ion channel recordings: Flickering event detection. *IEEE Trans. NanoBioscience*, 17(3):300–320, 2018.

[60] M. Pohlmann, F. Werner, and A. Munk. Minimax detection of localized signals in statistical inverse problems. *arXiv preprint, arXiv:2112.05648*, 2021.

[61] K. Proksch, F. Werner, and A. Munk. Multiscale scanning in inverse problems. *Ann. Statist.*, 46(6B):3569–3602, 2018.

[62] A. Shen and A. Volodin. Weak and strong laws of large numbers for arrays of rowwise END random variables and their applications. *Metrika*, 80(6-8):605–625, 2017.

[63] R. H. Shumway and D. S. Stoffer. *Time series analysis and its applications*. Springer Texts in Statistics. Springer-Verlag, New York, 2000.

[64] M. M. Siddiqui. On the inversion of the sample covariance matrix in a stationary autoregressive process. *Ann. Math. Statist.*, 29:585–588, 1958.

[65] D. Siegmund. *Sequential analysis*. Springer Series in Statistics. Springer-Verlag, New York, 1985.

[66] M. Singer, T. Krivobokova, and A. Munk. Kernel partial least squares for stationary data. *J. Mach. Learn. Res.*, 18:Paper No. 123, 41, 2017.

[67] S. H. Sung, S. Lisawadi, and A. Volodin. Weak laws of large numbers for arrays under a condition of uniform integrability. *J. Korean Math. Soc.*, 45(1):289–300, 2008.

[68] M. Talagrand. *The generic chaining*. Springer Monographs in Mathematics. Springer-Verlag, Berlin, 2005.

[69] A. B. Tsybakov. *Introduction to nonparametric estimation*. Springer Series in Statistics. Springer, New York, 2009.

[70] Y. Vardi, L. A. Shepp, and L. Kaufman. A statistical model for positron emission tomography. *J. Amer. Statist. Assoc.*, 80(389):8–37, 1985.

[71] X. Wang and S. Hu. Weak laws of large numbers for arrays of dependent random variables. *Stochastics*, 86(5):759–775, 2014.

[72] F. Werner and T. Hohage. Convergence rates in expectation for Tikhonov-type regularization of inverse problems with Poisson data. *Inverse Problems*, 28(10):104004, 15, 2012.

[73] Y. Yu. A review on minimax rates in change point detection and localisation. *arXiv preprint, arXiv:2011.01857*, 2020.

Markus Pohlmann

Curriculum Vitae

✉ *markus.pohlmann@mathematik.uni-goettingen.de*

Education

Since 2018 **Ph.D. Student**, *Georg-August-Universität Göttingen*,
Supervisors: Prof. Dr. Axel Munk and Prof. Dr. Frank Werner.

2018 **M.Sc. in Mathematics**, *Georg-August-Universität Göttingen*,
Thesis title: *Estimation in Semiparametric Mixture Models – from Algorithms to Asymptotics*,
Supervisor: Prof. Dr. Axel Munk.

2015 **B.Sc. in Mathematics**, *Georg-August-Universität Göttingen*,
Thesis title: *Gleichmäßige untere Schranken beim Kreisproblem*,
Supervisor: Prof. Dr. Valentin Blomer.

2011 **Abitur**, *Otto-Hahn-Gymnasium Springe*.

Publications and Preprints

[1] **Markus Pohlmann**, Frank Werner and Axel Munk. Minimax detection of localized signals in statistical inverse problems. *arXiv preprint, arXiv:2112.05648*, 2021.

[2] Farida Enikeeva, Axel Munk, **Markus Pohlmann** and Frank Werner. Bump detection in the presence of dependency: does it ease or does it load? *Bernoulli*, 26(4):3280-3310, 2020.

.

Conference Talks

2019 **32nd European Meeting of Statisticians**, *Palermo*.