INAUGURAL-DISSERTATION

zur
Erlangung der Doktorwürde
der
Naturwissenschaftlich-Mathematischen Gesamtfakultät
der Ruprecht-Karls-Universität
Heidelberg

vorgelegt von

Sandra Simone Schluttenhofer
M.Sc. Mathematik

aus Tübingen

Tag der mündlichen Prüfung:

Adaptive Minimax Testing for Inverse Problems

Betreuer: Prof. Dr. Jan Johannes

Acknowledgments

First of all, I am grateful to my supervisor Jan Johannes for his support, his advice and the many interesting scientific and non-scientific discussions we had. I want to thank my colleagues and friends at the universities in Heidelberg and Mannheim. I enjoyed our joint activities, all the game and movie nights, the hiking trips, the cooking experiments and the lively discussions during lunch very much.

In particular I would like to thank Alex for listening to all my doubts and troubles and always finding the right kind words of encouragement. I want to thank Stefan for the crafting time and the funny distractions. And finally, thank you, Christof, for helping me juggle everything. I also thank Dagmar Neubauer and Jutta Wiech for taking care of all the administrative issues and for the beautiful chilli plants that made my office even more suitable for fruitful discussions.

I am also deeply grateful to my friends and family who supported me through the many highs and lows during the last years. The Thursday night dinners, Monday night movies, Tuesday night card games, the midday walks and the many cups of coffee provided a great source of support, encouragement and happiness.

I had the pleasure of being a member of the Research Training Group (RTG) 1953 “Statistical Modeling of Complex Systems and Processes”, which provided an excellent working environment, and I thank the Deutsche Forschungsgemeinschaft (DFG) for the financial support. Moreover, I am grateful to the administration of the *Combined Faculty of Natural Sciences and Mathematics* of the Ruprecht-Karls-Universität Heidelberg.

I would like to offer special thanks to Fabienne Comte and Cristina Butucea for their inspiring suggestions during my visit in Paris last summer. Moreover, I have greatly benefited from the insightful discussions with Félix Beroud and Clément Marteau during their visit in Heidelberg and during my visit in Lyon.

Zusammenfassung

Die vorliegende Arbeit beschäftigt sich mit dem statistischen Testen für schlecht-gestellte inverse Probleme. Als Beobachtungen liegen lediglich verrauschte Versionen einer unbekannten Transformation der uns interessierenden Größe vor. Statistische Inferenz, bei der typischerweise eine Inversion der Transformation nötig ist, wird deshalb als inverses Problem bezeichnet. Besonders herausfordernd sind hierbei schlecht-gestellte inverse Probleme, bei denen die Inversion der Transformation instabil ist. Unsere vorgeschlagenen nicht-parametrischen Tests bewerten wir mittels eines nicht-asymptotischen Minimax-Kriteriums.

Die Arbeit besteht aus zwei Teilen, die sich mit unterschiedlichen schlecht-gestellten inversen Modellen beschäftigen. Im ersten Teil betrachten wir ein inverses Gaußsches Folgenmodell mit nur teilweise bekanntem Operator und im zweiten Teil die zirkuläre Faltung. In beiden Modellen leiten wir Minimax-Separationsradien her. Diese charakterisieren, wie weit ein Objekt von der Nullhypothese entfernt sein muss, damit der Unterschied von einem statistischen Test erkannt werden kann. Wir stellen zwei Testprozeduren vor. Zum einen betrachten wir einen (indirekten) Test, bei dem wir mit Hilfe eines Projektionsansatzes den Abstand zur Nullhypothese schätzen. Wir zeigen seine Minimax-Optimalität unter schwachen Annahmen. Der zweite (direkte) Test basiert stattdessen auf einer Schätzung des Abstandes im Bildraum des Operators und umgeht dadurch die Inversion des Operators. Auch hier charakterisieren wir die Situationen, in denen der Test minimax-optimal ist. Die Güte unserer Tests hängt – wie es in der nicht-parametrischen Statistik üblich ist – von der geeigneten Wahl eines Dimensionsparameters ab. Für die Optimalität des Tests wird Vorwissen über die zugrundeliegende Struktur benötigt. Deshalb beschäftigen wir uns außerdem mit adaptiven Teststrategien, die ohne solch ein Vorwissen auskommen. Wir wenden eine klassische Bonferroni-Aggregationsmethode auf unsere beiden Testprozeduren (direkt und indirekt) an und untersuchen die resultierenden Methoden auf ihre Optimalität. Verglichen mit den nicht-adaptiven Separationsradien stellen wir eine Verschlechterung um einen logarithmischen Faktor fest. Wir beweisen, dass dieser logarithmische Faktor ein unvermeidbarer Preis ist, der für Adaptivität bezahlt werden muss.

Unsere Testmethoden basieren auf der Schätzung eines quadratischen Funktionals, nämlich der Distanz zur Nullhypothese. Wir untersuchen den Zusammenhang zwischen den beiden statistischen Fragestellungen — dem Schätzen des quadratischen Funktionals und dem Testen – im zirkulären Faltungsmodell. Wir erläutern, wie sich Resultate für eines der beiden Probleme in den Kontext des anderen übertragen lassen. Abschließend betrachten wir Testprobleme im Zusammenhang mit Datenschutzbeschränkungen. Dabei werden die Daten, bevor sie zur statistischen Analyse zur Verfügung stehen, anonymisiert. Das heißt, die Daten werden vor der Weitergabe an die Statistikerin in einer bestimmten Art abgewandelt, um die Privatsphäre von Individuen zu schützen. Wir untersuchen, wie eine solche Privatisierung der Daten die Aussagekraft der statistischen Tests beeinflusst.

Abstract

This thesis deals with non-parametric hypothesis testing for ill-posed inverse problems, where optimality is measured in a non-asymptotic minimax sense. Loosely speaking, we observe only an approximation of a transformed version of the quantity of interest. Statistical inference, which usually requires an inversion of the transformation, is thus an inverse problem. Particularly challenging are ill-posed inverse problems, where the inverse transformation is not stable.

The thesis is divided into two parts, which investigate different ill-posed inverse models: the inverse Gaussian sequence space model with partially unknown operator and a circular convolution model. In both models we derive minimax separation radii of testing, which characterise how much an object has to differ from the null hypothesis to be detectable by a statistical test. We propose two types of testing procedures, an indirect and a direct one. The indirect test is based on a projection-type estimation of the distance to the null and we prove its minimax optimality under mild assumptions. The direct test is instead based on estimating the energy in the image space and thus avoids an inversion of the operator. We highlight the situations in which also the direct test performs optimally. As usual in non-parametric statistics, the performance of our tests depends on the optimal choice of a dimension parameter, which relies on prior knowledge of the underlying structure of the model. We derive adaptive testing strategies by applying a classical Bonferroni aggregation to both the direct and the indirect testing procedures and analyse their performance. Compared with the non-adaptive tests their radii face a deterioration by a log-factor, which we show to be an unavoidable cost to pay for adaptation.

Since our minimax optimal testing procedures are based on estimators of a quadratic functional, we further explore the connection between the two problems – quadratic functional estimation and minimax testing – in the circular convolution model. We show how results from one framework can be exploited in the other. Lastly, we consider minimax testing under privacy constraints, where the observations are deliberately transformed before being released to the statistician in order to protect the privacy of an individual.

Contents

Introduction **1**

I Inverse Gaussian sequence space model **11**

Inverse Gaussian sequence space model **13**

1 Adaptive minimax testing with partially known operators **19**
1.1 Introduction 19
1.2 Upper bound via an indirect testing procedure 24
1.3 Lower bound 29
1.4 Upper bound via a direct testing procedure 31
1.5 Adaptation 35
1.5.1 Description of the adaptation procedure 35
1.5.2 Adaptation to smoothness – indirect test 37
1.5.3 Adaptation to both smoothness and ill-posedness – direct test 42
1.6 Adaptive lower bound 45

A Auxiliary results **49**
A.1 Non-central χ^2-random variables 49
A.2 Balancing Lemma 50
A.3 Calculations for the χ^2-divergence 51
A.4 Mixtures 52

2 Testing of linear functionals **55**
2.1 Linear functional testing 55
2.2 Upper bound 57
2.3 Lower bound 60
2.4 Comparison to goodness-of-fit testing 61

B Auxiliary results **65**
B.1 Calculations for the χ^2-divergence 65

II Circular convolution 67

Circular convolution 69

3 Minimax testing and quadratic functional estimation 77
3.1 Introduction . . . 77
3.1.1 Quadratic functional estimation . . . 77
3.1.2 The testing task . . . 78
3.1.3 Methodology . . . 79
3.1.4 Minimax results . . . 80
3.2 Upper bound for the estimation risk . . . 82
3.3 Upper bound for the radius of testing . . . 87
3.4 Lower bound for the radius of testing . . . 90
3.5 Connection between quadratic functional estimation and testing . . . 92
3.6 Lower bound for the estimation risk . . . 93
3.7 Upper bound for the radius of testing via a direct test . . . 95

C Auxiliary results 101
C.1 Auxiliary results for proving lower bounds of testing . . . 101
C.2 Auxiliary results for proving lower bounds of estimation . . . 103

4 Adaptive minimax testing for circular convolution 105
4.1 Adaptive testing . . . 105
4.2 Upper bound via an indirect testing procedure . . . 108
4.3 Adaptive indirect testing procedure . . . 115
4.3.1 Aggregation of the indirect tests and the choice of the levels α_k . . . 116
4.3.2 Testing radius of the indirect max-test . . . 118
4.4 Upper bound via a direct testing procedure . . . 126
4.5 Adaptive direct testing procedure . . . 132
4.5.1 Aggregation of the direct tests and the choice of the levels α_k . . . 132
4.5.2 Testing radius of the direct max-test . . . 132
4.6 Adaptive lower bound . . . 140

D Auxiliary results 149
D.1 Preliminaries . . . 149
D.2 Auxiliary results used in the proof of Proposition 4.2.1 . . . 154
D.3 Auxiliary results used in the proof of Proposition 4.4.1 . . . 160
D.4 Calculations for the χ^2-divergence . . . 161

5 Testing under privacy constraints 165
5.1 Differential local privacy and privatized testing . . . 165
5.2 Upper bound for general privatization methods . . . 170
5.2.1 Upper bound via Laplace perturbation . . . 177
5.2.2 Upper bound via hypercube sampling . . . 182

Perspectives 187

Bibliography 190

Introduction

Inverse problems appear in many fields of science, for instance in climatology, economics and medicine. We consider non-asymptotic adaptive minimax testing for inverse problems. In this introduction we explain the main ideas for a general inverse model. The thesis then deals with the statistical investigations for two specific models; the inverse Gaussian sequence space model and the circular convolution model.

The testing task. Let $(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}})$ and $(\mathcal{G}, \langle\cdot,\cdot,\rangle_{\mathcal{G}})$ be two separable Hilbert spaces and let

$$T : \mathcal{H} \longrightarrow \mathcal{G}, h \longmapsto Th$$

be a linear operator. Based on a noisy observation of Th we aim to make inference on the element h. Specifically, for a given benchmark h° we consider the testing problem

$$H_0 : h = h^\circ \qquad \text{against} \qquad H_1 : h \neq h^\circ,$$

where optimality is measured in a minimax sense.

Inverse problems

Recovering the original element $h \in \mathcal{H}$ based on noisy versions of the image $g = Th \in \mathcal{G}$ is called **statistical inverse problem** since it typically requires an **inversion** of the operator T. Hadamard [1902] characterises a **well-posed** inverse problem, given by the equation $g = Th$, through three conditions.

- **Existence.** A solution exists, i.e. there exists an element $h \in \mathcal{H}$ such that $Th = g$.
- **Uniqueness.** The solution is unique, i.e. the operator T is injective.
- **Stability.** The inverse operator T^{-1} is continuous, i.e. for all $\varepsilon > 0$ there exists a $\delta > 0$ such that $\|\tilde{g} - g\|_{\mathcal{G}} \leq \delta$ implies $\|T^{-1}\tilde{g} - T^{-1}g\|_{\mathcal{H}} \leq \varepsilon$.

The first two conditions are minimal conditions for the meaningful recovery of the element of interest $h \in \mathcal{H}$. The third guarantees that a good approximation of g automatically yields a good approximation of $T^{-1}g = h$. In this thesis we consider **ill-posed** inverse problems, where the first two conditions are satisfied but the third condition is violated. If we, however, no longer require T^{-1} to be continuous, we cannot ensure a good approximation of h even if we can approximate g well. **Regularization methods** help to overcome this issue by replacing T^{-1} with a suitable continuous operator. We briefly mention two common strategies.

- **Tikhonov regularization.** Given an approximation $\tilde{g}$ of $g = Th$, instead of considering $T^{-1}\tilde{g}$, we take the minimiser of the Tikhonov functional

$$\mathcal{F} : \mathcal{H} \longrightarrow \mathbb{R}, f \longmapsto \mathcal{F}(f) := \|Tf - \tilde{g}\|_{\mathcal{G}}^2 + \lambda \|f\|_{\mathcal{H}}^2$$

for a regularization parameter $\lambda > 0$. That is, as an approximation of the solution h we consider

$$\tilde{h} \in \arg\min_{f \in \mathcal{H}} \mathcal{F}(f).$$

The term $\lambda \|f\|_{\mathcal{H}}^2$ in the Tikhonov functional $\mathcal{F}(f)$ is a regularization term, which enforces solutions of smaller norm. If $\|\tilde{g} - g\|_{\mathcal{G}} \longrightarrow 0$ and the regularization parameter is appropriately chosen, the approximation $\tilde{h}$ converges to the solution h. We refer to Chapter 5 in Engl et al. [1996] (e.g. Theorem 5.2) for more details.

- **Projection regularization.** Let $(b_j)_j$ be a basis of the Hilbert space $\mathcal{H}$ and denote by $\mathcal{H}_k := \mathrm{Lin}(b_j, j \le k)$ the space spanned by the first k basis elements. Projection regularization searches for the best approximation contained in the finite dimensional subspace, i.e. given an approximation $\tilde{g}$ of $g = Th$ we consider a minimiser in $\mathcal{H}_k$ of the functional

$$\mathcal{F} : \mathcal{H} \longrightarrow \mathbb{R}, f \longmapsto \mathcal{F}(f) := \|Tf - \tilde{g}\|_{\mathcal{G}}^2$$

i.e. an element

$$\tilde{h} \in \arg\min_{f \in \mathcal{H}_k} \mathcal{F}(f).$$

as an approximation of the solution h. In this case the dimension k is the regularization parameter. We refer to Chapter 3.3 in Engl et al. [1996] for more details and conditions under which the approximation $\tilde{h}$ converges to the true solution h provided that $\tilde{g}$ approaches g.

Example (Deconvolution). Let us illustrate the concepts with a **deconvolution** model. We consider the Hilbert space $\mathcal{H} = \mathcal{G} = \mathscr{L}^2_{\mathrm{per}}$ of 1-periodic complex-valued functions defined on $\mathbb{R}$, which are square integrable on $[0, 1)$. The space $\mathscr{L}^2_{\mathrm{per}}$ is equipped with its usual inner product $\langle f, g\rangle_{\mathscr{L}^2_{\mathrm{per}}} = \int_0^1 f(x)\overline{g(x)}\mathrm{d}x$. For $\xi \in \mathscr{L}^2_{\mathrm{per}}$ we define the convolution operator

$$\begin{aligned} T_\xi : \mathscr{L}^2_{\mathrm{per}}[0,1) &\longrightarrow \mathscr{L}^2_{\mathrm{per}}[0,1), \\ h &\longmapsto T_\xi h \end{aligned}$$

given by

$$T_\xi h(y) := \int_0^1 \xi(y - x)h(x)\mathrm{d}x, \qquad \text{for} \qquad y \in [0, 1).$$

Denote by $(\boldsymbol{e}_j)_{j\in\mathbb{Z}}$ with $\boldsymbol{e}_j(x) = \exp(2\pi i j x)$, $x \in [0, 1)$ the Fourier basis of $\mathscr{L}^2_{\mathrm{per}}$ and by $h_j := \langle h, \boldsymbol{e}_j\rangle_{\mathscr{L}^2_{\mathrm{per}}} \in \mathbb{C}$ the j-th Fourier coefficient of an element $h \in \mathscr{L}^2_{\mathrm{per}}$. By Parseval's theorem an element h belongs to $\mathscr{L}^2_{\mathrm{per}}$ if and only if its Fourier coefficients are square summable. The convolution operator has a representation in terms of Fourier coefficients given by

$$T_\xi h = \sum_{j\in\mathbb{Z}} \xi_j h_j \boldsymbol{e}_j. \tag{$*$}$$

Let us investigate Hadamard's conditions for well-posedness.

- **Existence.** The representation $(*)$ implies that any combination of g, h and $\xi \in \mathscr{L}^2_{\mathrm{per}}$ with $g = T_\xi h$ satisfies $g_j = \xi_j h_j$ for all $j \in \mathbb{Z}$. Hence, for a given pair g and ξ there exists a solution if and only if the sequence $(\frac{g_j}{\xi_j}\mathbb{1}_{\{\xi_j\neq 0\}})_{j\in\mathbb{Z}}$ is square summable. In this case a solution h of $T_\xi h = g$ is given through its Fourier coefficients $(h_j)_{j\in\mathbb{Z}} = (\frac{g_j}{\xi_j}\mathbb{1}_{\{\xi_j\neq 0\}})_{j\in\mathbb{Z}}$.

- **Uniqueness.** Injectivity of the operator T_ξ can be expressed in terms of the Fourier coefficients of the convolution function ξ, it requires $\xi_j \neq 0$ for all $j \in \mathbb{Z}$.

- **Stability.** Naturally, if $\xi_j \neq 0$ for all $j \in \mathbb{Z}$, we can represent the inverse of T_ξ on its natural domain as

$$T_\xi^{-1} g = \sum_{j\in\mathbb{Z}} \frac{1}{\xi_j} g_j \boldsymbol{e}_j.$$

The inverse operator T_ξ^{-1} cannot be continuous, since $\xi \in \mathscr{L}^2_{\text{per}}$ implies that its Fourier coefficients tend to zero $|\xi_j| \longrightarrow 0$ for $|j| \longrightarrow \infty$. Indeed, the Fourier basis $(\boldsymbol{e}_n)_{n\in\mathbb{Z}}$ satisfies $\|\boldsymbol{e}_n\|^2_{\mathscr{L}^2_{\text{per}}} = 1$ for all $n \in \mathbb{Z}$ due to the orthonormality, but also

$$\left\|T_\xi^{-1}\boldsymbol{e}_n\right\|^2_{\mathscr{L}^2_{\text{per}}} = \frac{1}{|\xi_n|^2} \longrightarrow \infty \qquad \text{for } n \to \infty.$$

Thus, the operator is unbounded and, hence, not continuous.

Summarizing, **deconvolution** is an ill-posed inverse problem. The main issue is the discontinuity of the deconvolution operation. Let us examine the two regularization methods in this specific situation.

- **Tikhonov regularization.** For the deconvolution model the Tikhonov regularization, which is generally given implicitly as a minimization task, has an explicit representation (see Engl et al. [1996], Chapter 5.1.) given by

$$\tilde{h} = \sum_{j\in\mathbb{Z}} \frac{\overline{\xi_j}}{|\xi_j|^2 + \lambda} \tilde{g}_j \boldsymbol{e}_j$$

for an approximation $\tilde{g}$ of $g = Th$. The approximation error is now no longer amplified through the multiplication with the unbounded coefficients $\frac{1}{\xi_j}$, $j \in \mathbb{Z}$, which have instead been replaced by bounded counterparts $\frac{\overline{\xi_j}}{|\xi_j|^2+\lambda}$, $j \in \mathbb{Z}$.

- **Projection regularization.** Also the projection regularization takes a simpler form in the deconvolution model, it is given by

$$\tilde{h} = \sum_{j=-k}^{k} \frac{\tilde{g}_j}{\xi_j} \boldsymbol{e}_j,$$

where the truncation to only a finite number of dimensions stabilizes the regularized solution.

Non-asymptotic minimax testing theory.

Basic notions from statistical testing. Denote by $\mathcal{Y}$ the space of observations and by $\mathbb{P}_h$ the probability distribution associated with noisy observations of Th. A statistical test is a measurable function $\Delta : \mathcal{Y} \longrightarrow \{0, 1\}$. By convention, for a observation $y \in \mathcal{Y}$ we understand $\Delta(y) = 1$ as the decision to reject the null hypothesis and $\Delta(y) = 0$ as accepting the null. There are two kinds of errors that can occur. The **type I error** is to reject the null although it is true, this occurs with probability $\mathbb{P}_{h^\circ}(\Delta = 1)$. The **type II error** arises when accepting the null although it is not true, i.e. for some h contained in the alternative H_1 it happens with probability $\mathbb{P}_h(\Delta = 1)$. For $\alpha \in (0,1)$ a test Δ is said to have **level** α, if the type I error probability is bounded by α, i.e. $\mathbb{P}_{h^\circ}(\Delta = 1) \leq \alpha$. It is said to be **(1-$\beta$)-powerful**, $\beta \in (0,1)$, for the element $h \in \mathcal{H}$ if the type II error probability is bounded by β, i.e. $\mathbb{P}_h(\Delta = 0) \leq \beta$.

Minimax theory. Given a test $\Delta : \mathcal{Y} \longrightarrow \{0,1\}$ we measure its performance by how well it is able to distinguish between the null hypothesis and elements that are in some sense separated from the null. Formally, for a separation radius $\rho > 0$ and a non-parametric regularity class $\mathcal{E}$ we consider the testing task

$$H_0 : h = h^\circ \qquad \text{against} \qquad H_1^\rho : \|h - h^\circ\|_{\mathcal{H}} \geq \rho, h - h^\circ \in \mathcal{E}, \tag{ITT}$$

where the null hypothesis H_0 and the alternative H_1^ρ are separated to make them statistically distinguishable. If h° is the null element in the Hilbert space, the testing task (ITT) is called **signal detection**, since the aim is to detect a non-zero signal h. Otherwise we call it a **goodness-of-fit** testing task. We define the **maximal risk** of a test

$$\mathcal{R}\left(\Delta \mid \mathcal{E}, \rho\right) := \mathbb{P}_{h^\circ}\left(\Delta = 1\right) + \sup_{\substack{\|h-h^\circ\|_{\mathcal{H}} \geq \rho \\ h-h^\circ \in \mathcal{E}}} \mathbb{P}_h\left(\Delta = 0\right),$$

as the sum of the type I error probability and the maximal type II error probability over the ρ-separated alternative. The difficulty of the testing task is then characterised by the **minimax risk**

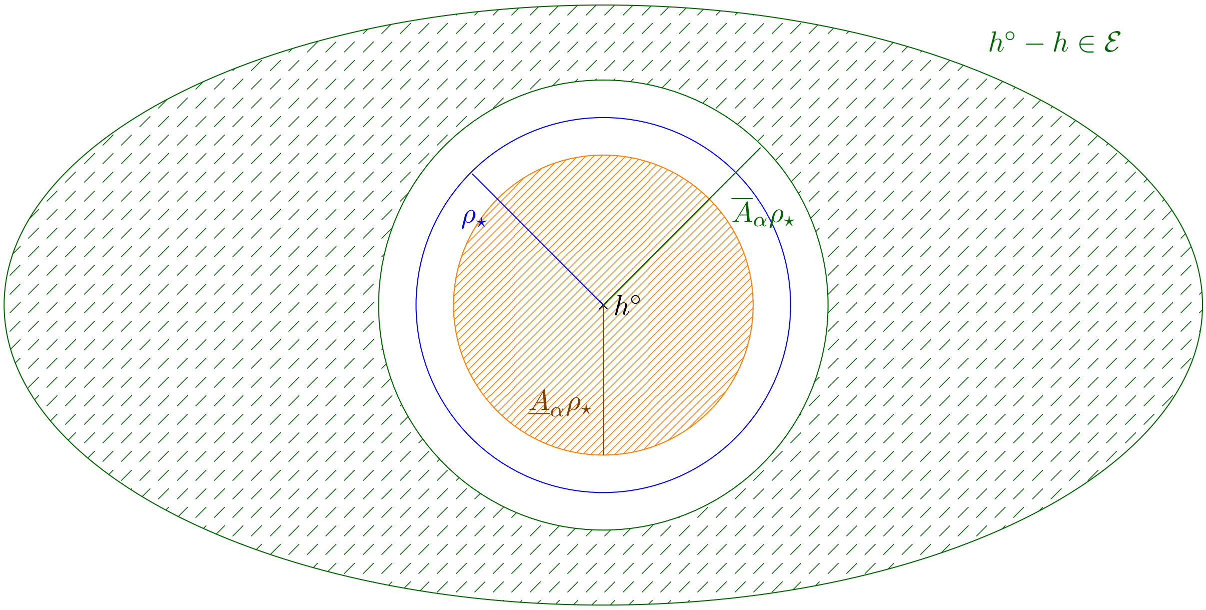
$$\mathcal{R}\left(\mathcal{E}, \rho\right) := \inf_{\Delta} \mathcal{R}\left(\Delta \mid \mathcal{E}, \rho\right),$$

where the infimum is taken over all possible tests. Generally three factors influence the difficulty of the testing task: the ill-posedness of the operator T, the regularity class $\mathcal{E}$, and the noise level or the number of observations with which we observe $Th = g$. In particular, we are interested in the minimal distance by which we need to separate the null hypothesis and the alternative such that there exists a test that can tell them apart with high probability. Formally, a radius $\rho_\star^2$ is called **minimax separation radius** or **minimax radius of testing** if for all $\alpha \in (0,1)$ there exist constants $\underline{A}_\alpha, \overline{A}_\alpha > 0$ such that

(i) for all $A \geq \overline{A}_\alpha$ we have $\mathcal{R}\left(\mathcal{E}, A\rho_\star\right) \leq \alpha$, (upper bound)

(ii) for all $A \leq \underline{A}_\alpha$ we have $\mathcal{R}\left(\mathcal{E}, A\rho_\star\right) \geq 1 - \alpha$. (lower bound)

Condition (i) essentially states that if we separate the null and the alternative further than $\overline{A}_\alpha \rho_\star$, the minimax risk is smaller than α. We show the upper bound (i) by constructing a testing procedure $(\Delta_\alpha)_{\alpha \in (0,1)}$, which satisfies the required risk bound. Condition (ii) guarantees the opposite: if we allow elements that are closer to the null hypothesis than $\underline{A}_\alpha \rho_\star$ the sum of error probabilities is large, no matter which test we choose. The lower bound (ii) is typically proved by exploiting reduction arguments, showing the lower risk bound for an arbitrary test.



Visualization of the minimax radius of testing. Elements in the green-striped area are far enough away from the null, i.e. separated by at least $\overline{A}_\alpha \rho_\star$, to be detected with high probability. That is, there exists a test with minimax risk over the $\overline{A}_\alpha \rho_\star$-separated alternative smaller than α. Elements closer to the null hypothesis h° than $\underline{A}_\alpha \rho_\star$ (orange-striped area) cannot be statistically distinguished from the null. If we include such elements in the alternative, the risk of any test is larger than $1 - \alpha$. The *detection boundary* is given by the minimax separation radius $\rho_\star$.

Other notions of separation radii. The notion of the minimax separation radius used in this thesis can e.g. be found in Collier et al. [2017]. There exists an alternative definition, which was pioneered by Baraud [2002]. Therein the levels $\alpha \in (0, 1)$ and $\beta \in (0, 1)$ for the error probabilities are fixed and one searches for the smallest separation radius such that there exists an α-test that is $(1-\beta)$-powerful over all separated alternatives. Since the derivations of our upper bounds are done by controlling the type I and the maximal type II error probabilities separately (compare Proposition 1.2.1, Proposition 2.2.1 and Proposition 4.2.1), it would also be possible to derive separation radii in the sense of Baraud [2002]. We emphasise that both definitions of the minimax radius of testing are entirely non-asymptotic, in other words they do not require the noise level to vanish or the number of observations to tend to infinity. Nonetheless, often one is interested in the behaviour of the separation radii when the noise level tends to zero or the sample size increases. In this situation the radii are referred to as **rates of testing** or **separation rates**. Analogously to the estimation theory for inverse problems, where there exist asymptotic (cp. Carrasco et al. [2007]) as well as non-asymptotic approaches (cp. Cavalier [2008] or Efromovich and Koltchinskii [2001]), there is an asymptotic framework for minimax testing, which was mostly established in the series of papers Ingster [1993a], Ingster [1993b], Ingster [1993c]. We refer to the monograph Ingster and Suslina [2012] for an extensive overview and the paper Marteau and Sapatinas [2015], which explores the connection between the asymptotic and non-asymptotic setting.

> The first objective of this thesis is to derive (non-asymptotic) minimax separation radii for general classes of non-parametric alternatives $\mathcal{E}$ and operators T. For the upper bounds we construct tests that are based on projection-type estimators $\hat{q}_k^2$ of $\|h - h^\circ\|_{\mathcal{H}}^2$, where $k \in \mathbb{N}$ is a truncation parameter. This is a natural approach when considering the testing problem

(ITT) since the hypotheses are separated by this quantity. We reject the null hypothesis if the estimated value $\hat{q}_k^2$ exceeds a certain threshold. Deriving upper bounds then requires bounds of the quantiles of these projection-type estimators under the null hypothesis and the alternatives. The upper bounds mimic a classical bias^2 - variance trade off, which calls for an optimal choice of the truncation parameter k. For the upper bounds the essential tools (i.e. Bernstein-type concentration inequalities) highly depend on the observational model and the particular Hilbert space that is considered. Lower bounds are shown by reducing the testing risk to a distance of probability measures. The key points in these proofs are the construction of suitable so-called candidate elements and bounds for the χ^2-divergence or the Kullback-Leibler divergence over mixtures of these candidate elements.

Since our minimax optimal tests are based on estimators of the quadratic functional $h \mapsto \|h - h^\circ\|_{\mathcal{H}}^2$, another aim of this thesis is to highlight the connection between the two problems – quadratic functional estimation and minimax testing. We show how results from one framework can be exploited in the other. Although the problems are clearly closely connected, they feature structurally different behaviour in the radii and we come to the conclusion that, roughly speaking, testing is always faster than estimation.

Direct and indirect approaches. There occurs an interesting phenomenon when treating minimax testing in inverse models. Generally speaking, the ill-posedness of an inversion causes additional difficulties, which can be avoided in the context of testing. Given noisy observations of Th it is natural to consider a **direct** testing task

$$H_0^{\mathrm{D}} : Th = Th^\circ \qquad \text{against} \qquad H_1^{\mathrm{D}} : Th \neq Th^\circ,$$

which can be solved without an inversion of the operator T. By investigating the direct testing problem we shift the statistical inference from the pre-image space $\mathcal{H}$ to the image space $\mathcal{G}$ of the operator T. Moreover, for an injective (known) operator T the null hypotheses of the direct problem ($Th = Th^\circ$) and of the indirect problem ($h = h^\circ$) coincide. By using a test that is conceptionally constructed to solve the **direct** testing problem, i.e. where the statistical analysis is conducted in the space $\mathcal{G}$, for the **indirect** testing problem, we circumvent the additional instability typical for inverse problems. Formulating the direct testing problem from a minimax point of view, we again introduce a separation of the null hypothesis and the alternative, which yields

$$H_0^{\mathrm{D}} : Th = Th^\circ \qquad \text{against} \qquad H_1^{\mathrm{D},\rho} : \|Th - Th^\circ\|_{\mathcal{H}} \geq \tilde{\rho}, Th - Th^\circ \in \mathcal{F}, \tag{DTT}$$

for an appropriately transformed regularity class $\mathcal{F}$. Laurent et al. [2012] show that in specific situations a test that is minimax optimal for direct testing task (DTT) is also minimax optimal for indirect testing task (ITT). In these situations an inversion of the operator is unnecessary and, thus, should be avoided. The key point of such a result is an embedding of the form

$$\{\|h - h^\circ\|_{\mathcal{H}} \geq \rho, h - h^\circ \in \mathcal{E}\} \subseteq \{\|Th - Th^\circ\|_{\mathcal{H}} \geq \tilde{\rho}, Th - Th^\circ \in \mathcal{F}\},$$

which naturally involves both characteristics of the operator T and the regularity classes $\mathcal{E}$ and $\mathcal{F}$.

The second objective of this thesis is to investigate the performance of direct testing procedures (constructed to solve (DTT)) for the indirect testing task (ITT) and to derive conditions under which they perform optimally. That is, we aim to characterise the situations in which an inversion is advisable and the situations in which it can be avoided in terms of properties of the regularity class $\mathcal{E}$ and the ill-posedness of the operator T. In particular, we explore what happens in the case of an **unknown operator** T. We point out that the null hypothesis Th° of the direct testing problem is — in the case of unknown T — only prespecified if h° is the null element in the Hilbert space $\mathcal{H}$ (since it gets mapped to the null element in the Hilbert space $\mathcal{G}$ by the linear operator T). Therefore, there is a natural distinction between signal detection (testing against the null element) and goodness-of-fit problems (testing against a prescribed non-zero element) if the operator is unknown. In signal detection we build tests that are based on a projection-type estimation of $\|Th\|_{\mathcal{H}}^2$, where we have direct access to noisy observations of Th. For the goodness-of-fit it is natural to base a test on the estimation of $\|Th - Th^\circ\|_{\mathcal{H}}^2$. However, if T is unknown so is Th°, which thus has to be estimated. Nevertheless, our upper bounds for the direct test are again given by a bias^2 - variance trade off and in many cases coincide with the minimax separation radius (up to constants).

Adaptation. The tests constructed for solving (ITT) and (DTT) typically requiring an optimal choice of a tuning parameter depending on properties of the regularity class $\mathcal{E}$ and the operator T. Hence, these testing procedures are not **adaptive**, i.e. not **assumption-free**. Preferably we want our tests to perform (nearly) optimal over a wide range of classes $\mathcal{E}$ (and operators T) simultaneously and without requiring any a priori knowledge about the underlying structure. In non-parametric estimation adaptation methods often include data-driven choices of tuning parameters (cp. Birgé [2001]) or aggregation approaches (cp. Tsybakov [2004]). In nonparametric testing, adaptation is most commonly approached by multiple testing procedures. A classical method is the Bonferroni aggregation of tests, where one considers a *maximum*-test over an appropriately chosen (finite) class, which rejects the null as soon as one of the tests in the collection does. Let us be more precise. Assume we have constructed a finite collection of tests $(\Delta_{k,\alpha_k})_{k\in\mathcal{K}}$, where for each tuning parameter $k \in \mathcal{K}$ the test $\Delta_{k,\alpha_k} = \mathbb{1}_{\{\hat{q}_k^2 > \tau_k(\alpha_k)\}}$ with test statistic $\hat{q}_k^2$ and threshold $\tau_k(\alpha_k)$ is of level $\alpha_k \in (0,1)$. Let $\sum_{k\in\mathcal{K}} \alpha_k = \alpha$ and consider the *max*-test

$$\Delta_{\mathcal{K},\alpha} = \mathbb{1}_{\left\{\max\limits_{k\in\mathcal{K}}\left\{\hat{q}_k^2 - \tau_k(\alpha_k)\right\} > 0\right\}}.$$

The maximum structure of the test allows for an easy control of the type I error probability,

$$\mathbb{P}_{h^\circ}\left(\Delta_{\mathcal{K},\alpha} = 1\right) \leq \sum_{k\in\mathcal{K}} \mathbb{P}_{h^\circ}\left(\Delta_{k,\alpha_k} = 1\right) = \sum_{k\in\mathcal{K}} \alpha_k = \alpha$$

and the type II error probability

$$\mathbb{P}_h\left(\Delta_{\mathcal{K},\alpha} = 0\right) \leq \min_{k\in\mathcal{K}} \mathbb{P}_h\left(\Delta_{k,\alpha_k} = 0\right).$$

Therefore, the max-test behaves (almost) as well as the best test contained in the collection. Hence, for each alternative that we want to adapt to, the set $\mathcal{K}$ should be chosen such that there exists a (nearly) optimal test in the collection. The notable difference is that α has been replaced with α_k and the cost to pay for adaptation is characterised by the difference of the power of the tests Δ_{k,α_k} and $\Delta_{k,\alpha}$. Thus, aggregation (and therefore adaptation) of tests typically involves a deterioration of the radius of testing. This idea is formalized by introducing an **adaptive factor**, which characterises the loss. A natural question arises: Is the cost that we pay for adaptation due to a suboptimal testing strategy or is it caused intrinsically by the problem and is, hence, unavoidable?

The third objective of this thesis is to propose adaptive testing procedures, to characterise the cost to pay for adaptation and to show that this cost is unavoidable. Though the aggregation of tests can be executed for many testing strategies, determining minimal (unavoidable) adaptive factors is especially demanding. Typically it requires the construction of α-level tests such that their dependence on α is in some sense optimal (often this means it mimics the behaviour of Gaussian quantiles). In the notation above this translates to minimal changes in the threshold $\tau_k(\alpha)$ when α is replaced with the smaller quantity α_k. It requires a sharp control of the quantiles of a test statistic using for example Bernstein-type inequalities. We characterise adaptive factors in terms of arbitrary collections $\mathcal{K}$ and general sets of alternatives. Unavoidability of a deterioration in the rate for adaptation has mainly been considered in the setting of asymptotic minimax rates and only for specific smoothness classes (Spokoiny [1996]). We provide general conditions on the complexity of the set of alternatives and the size of the unavoidable adaptive factor. Interestingly, this general result allows us to derive the minimal adaptive factors for widely considered sets of alternatives. Lower bounds for the adaptive factors are particularly challenging since they involve an additional mixture over various alternatives, which complicates the calculations e.g. for χ^2-divergences over mixtures and requires a delicate construction of the candidate densities.

Privacy constraints. In recent years making data publicly available while still protecting the privacy of an individual has become an increasingly important task. Statistical inference under a local differential privacy constraint, where only anonymized versions of the observations are available, is a highly demanding challenge. Roughly speaking, each data holder transforms the observation according to a random mechanism before passing the data on to the statistician. Ideally such a mechanism should protect the privacy of an individual while preserving the information that is needed to make meaningful statistical inference. In general, the task here is twofold: one has to develop data-release mechanisms that guarantee a certain level of privacy and statistical methods that perform well based on privatized observations.

We consider minimax testing for inverse problems in a local differential privacy setting and investigate how different levels of privacy protection influence the separation radii.

Contributions and structure of this thesis

In this thesis we consider two statistical models: The Inverse Gaussian sequence space model (part I) and Circular convolution (part II), which are introduced in detail at the beginning of the respective parts of this thesis. Let us briefly put the two models in the context of the general inverse problem described in this introduction. In the inverse Gaussian sequence space model we consider the Hilbert spaces $\mathcal{H} = \mathcal{G} = \ell^2(\mathbb{N})$ of square-summable real-valued sequences and a multiplication operator $T_\lambda((x_j)_{j\in\mathbb{Z}}) = (\lambda_j x_j)_{j\in\mathbb{Z}}$. We make inference on the sequence $(x_j)_{j\in\mathbb{Z}}$ based on an observation of the image sequence $(\lambda_j x_j)_{j\in\mathbb{Z}}$ contaminated by additive Gaussian noise. In the circular convolution model we consider Hilbert spaces $\mathcal{H} = \mathcal{G} \subseteq \mathscr{L}^2[0,1)$ of square-integrable densities and a (circular) convolution operator $T_\xi f = f \circledast \xi$. We assume that we have independent and identically distributed observations from $T_\xi f$ at our disposal and aim to make inference on f. In each chapter we mention further references regarding the respective models and specific testing tasks in the **related literature**-paragraphs, putting the contributions of this thesis into more detailed context. Let us summarize the main results of this thesis.

Inverse Gaussian sequence space model

- Adaptive minimax testing with partially known operators (Chapter 1)
  In this chapter we derive minimax separation radii, compare direct and indirect testing procedures and carry out adaptation strategies for both. The main challenge of this chapter is the fact that the multiplication operator is unknown, but can be observed contaminated by additive noise. Interestingly, we show that for unknown operators the minimax radii depend on the null hypothesis h° and that in the case of signal detection the error in the

operator does not appear in the radii. The results of this chapter are published in the preprint Schluttenhofer and Johannes [2020a].

- Testing of linear functionals (Chapter 2)
  In this chapter instead of testing h° we test the value of a linear functional $L(h^\circ)$ and derive the corresponding minimax radii of testing. The content of this chapter originated from a discussion with Félix Beroud and Clément Marteau from the University of Lyon I.

Circular convolution model

- Minimax testing and quadratic functional estimation (Chapter 3)
  In this chapter we explore the connection between (minimax) quadratic functional estimation and (minimax) testing, which is natural since in the testing task the null hypothesis and the alternative are separated by a quadratic functional $h \mapsto \|h - h^\circ\|^2_{\mathcal{H}}$. We derive both minimax radii of testing and minimax estimation rates. We show that – although the problems are clearly closely connected – the typical elbow effect that occurs in quadratic functional estimation does not appear in the testing task. Our proofs yield a heuristic explanation for this effect: the elbow effect is caused by elements with large energy (i.e. large $\|h - h^\circ\|^2_{\mathcal{H}}$), these are difficult to estimate, but easy to test, since they are far away from the null hypothesis. The results of this chapter are published in the preprint Schluttenhofer and Johannes [2020b].

- Adaptive minimax testing for circular convolution (Chapter 4)
  In this chapter we again derive minimax separation radii, compare direct and indirect testing procedures and carry out adaptation strategies for both. Considering minimax testing in a circular convolution model is a natural extension of the Gaussian sequence space model, since we identify functions on the circle with their coefficients (forming a sequence) with respect to an appropriately chosen basis. There are, however, some new challenges. Firstly, the objects we consider are no longer Gaussian, which requires the application of different concentration results to prove upper bounds. Moreover, concerning lower bounds, which involve controlling χ^2-divergences between mixtures over candidate densities, we require estimates also for non-Gaussian distributions. The results of this chapter are published in the preprint Schluttenhofer and Johannes [2020c].

- Testing under privacy constraints (Chapter 5)
  In this chapter we assume that only privatized samples, i.e. deliberately transformed observations, are available and investigate privatization mechanisms and corresponding testing strategies. The first insight of this chapter is the fact that it is not advisable to perturb the observations themselves since this mimics an artificial convolution, which results in a significant deterioration of the rates. Instead perturbation of the coefficients of the density of interest in a certain basis are considered. We establish a general upper bound which can be applied to an arbitrary privatization method of the coefficients and provides bounds involving the variance of the privatized coefficients. We show that the standard privatization method of adding Laplace noise fails to perform optimally and is outperformed by a more involved privatization method called hypercube sampling. Comparing our results with findings in direct models, we conjecture our obtained radii to be minimax optimal and provide a preliminary framework for a possible lower bound in the Perspectives section.

Part I

Inverse Gaussian sequence space model

Inverse Gaussian sequence space model

Let $(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}})$ and $(\mathcal{G}, \langle\cdot,\cdot,\rangle_{\mathcal{G}})$ be separable infinite-dimensional real Hilbert spaces[1] and let

$$T : \mathcal{H} \longrightarrow \mathcal{G}, h \longmapsto Th$$

be a linear bounded[2] operator. In this section we specify the kind of noisy observations from Th that we consider in the first part of this thesis. We assume that we can observe the element Th projected onto *test functions*, i.e. the observable quantities are of the form $\langle Th, f\rangle_{\mathcal{G}}$ for $f \in \mathcal{G}$ and are contaminated by additive Gaussian noise. Firstly, we explain how to translate observable quantities of the form $\langle Th, f\rangle_{\mathcal{G}}$ for $f \in \mathcal{G}$ into an (indirect) sequence space model and give some examples where this occurs naturally. Next, we give a motivation for considering additive Gaussian errors in the sequence model context.

Sequence models

We distinguish three cases for the operator T. Let us start with the elementary case.

- Let $(\mathcal{H}, \langle\cdot,\cdot\rangle_{\mathcal{H}}) = (\mathcal{G}, \langle\cdot,\cdot,\rangle_{\mathcal{G}})$ and let $T = \mathrm{Id}_{\mathcal{H}}$ be the identity on $\mathcal{H}$.

Let $(\varphi_j)_{j\in\mathbb{N}}$ be an orthonormal basis of $\mathcal{H}$. Such a countable basis exists, since $\mathcal{H}$ is assumed to be separable. For any $j \in \mathbb{N}$ and $h \in \mathcal{H}$ let us define

$$\langle Th, \varphi_j\rangle_{\mathcal{H}} = \langle h, \varphi_j\rangle_{\mathcal{H}} =: \theta_j,$$

where the sequence of coefficients $(\theta_j)_{j\in\mathbb{N}}$ characterises the element of interest completely through the basis representation $h = \sum_{j\in\mathbb{N}}\langle h, \varphi_j\rangle_{\mathcal{H}}\varphi_j = \sum_{j\in\mathbb{N}}\theta_j\varphi_j$. The observational model is then given by the sequence

$$Y_j = \theta_j + \varepsilon\xi_j \qquad \text{with} \qquad \xi_j \overset{\text{iid}}{\sim} \mathcal{N}(0,1), \qquad j \in \mathbb{N}, \tag{GSSM}$$

which is called the **(direct) Gaussian sequence space model** with noise level $\varepsilon > 0$. It is called *direct* since we have direct access to noisy observations of the sequence of interest $(\theta_j)_{j\in\mathbb{N}}$. The model GSSM has received a lot of attention in the statistics literature, we only mention Baraud [2002], Ingster and Suslina [2012], Collier et al. [2017] and refer to the references therein.

- Let T be an injective operator with known singular value decomposition.

[1] A Hilbert space is a real or complex-valued complete metric space with a distance function that is induced by an inner product. Complete means that every Cauchy sequence has a limit within the space. A Hilbert space is separable if and only if it has a countable orthonormal basis.

[2] T is bounded if and only if there exists an $M > 0$ such that $\|Th\|_{\mathcal{G}} \leq M\,\|h\|_{\mathcal{H}}$ for all $h \in \mathcal{H}$.

We call a triple $(\varphi_j, \psi_j, \lambda_j)_{j\in\mathbb{N}}$, consisting of a basis $(\varphi_j)_{j\in\mathbb{N}}$ of $\mathcal{H}$, an orthonormal system $(\psi_j)_{j\in\mathbb{N}}$ in $\mathcal{G}$ and a sequence $(\lambda_j)_{j\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$, a **singular value decomposition** of the operator T if it can be represented in the form

$$Th = \sum_{j\in\mathbb{N}} \lambda_j \langle h, \varphi_j\rangle_{\mathcal{H}} \psi_j \qquad \text{for all } h \in \mathcal{H}.$$

The elements $(\varphi_j)_{j\in\mathbb{N}}$ and $(\psi_j)_{j\in\mathbb{N}}$ are called **singular vectors**, $(\lambda_j)_{j\in\mathbb{N}}$ are called **singular values**. If the singular value decomposition of an operator is known, we can project the element $Th \in \mathcal{G}$ onto the singular vectors. For any $j \in \mathbb{N}$ and $h \in \mathcal{G}$ we then obtain

$$\langle Th, \psi_j\rangle_{\mathcal{G}} = \langle \sum_{k\in\mathbb{N}} \lambda_k \langle h, \varphi_k\rangle_{\mathcal{H}} \psi_k, \psi_j\rangle_{\mathcal{G}} = \lambda_j \langle h, \varphi_j\rangle_{\mathcal{H}} =: \lambda_j\theta_j,$$

where again the sequence of coefficients $(\theta_j)_{j\in\mathbb{N}}$ characterises the element of interest completely through the basis representation $h = \sum_{j\in\mathbb{N}} \langle h, \varphi_j\rangle_{\mathcal{H}}\varphi_j = \sum_{j\in\mathbb{N}} \theta_j\varphi_j$. The observational model is given by

$$Y_j = \lambda_j\theta_j + \varepsilon\xi_j \qquad \text{with} \qquad \xi_j \overset{\text{iid}}{\sim} \mathcal{N}(0,1), \qquad j \in \mathbb{N}, \tag{IGSSM}$$

which is called the **indirect Gaussian sequence space model** with noise level $\varepsilon > 0$. Since T is assumed to be injective, we have $\lambda_j > 0$ for all $j \in \mathbb{N}$ and the parameter $(\theta_j)_{j\in\mathbb{N}}$ is identifiable in the model (IGSSM). If $\lambda_j \longrightarrow 0$ for $j \longrightarrow \infty$, the model (IGSSM) is called **ill-posed**, since a decaying sequence $(\lambda_j)_{j\in\mathbb{N}}$ weakens the signal of interest $(\theta_j)_{j\in\mathbb{N}}$ and, thus, inference on $(\theta_j)_{j\in\mathbb{N}}$ becomes more difficult. Otherwise it is called **well-posed**. The degree of ill-posedness is measured by the decay of $(\lambda_j)_{j\in\mathbb{N}}$. The problem is called **mildly ill-posed** if the sequence $(\lambda_j)_{j\in\mathbb{N}}$ decays polynomially and **severely ill-posed** if it decays at an exponential rate. A singular value decomposition with decaying singular values (yielding an ill-posed model) exists for instance for compact operators (Werner [2006], Theorem VI.3.6.). The model (IGSSM) is for example investigated in Cavalier and Tsybakov [2002], Ermakov [2006], Cavalier [2008] (all three considering estimation of $(\theta_j)_{j\in\mathbb{N}}$), Laurent et al. [2012], Laurent et al. [2011], Ingster et al. [2012a], Ingster et al. [2012b] (considering testing tasks).

► Let T be an injective operator with known singular vectors (but unknown singular values).

In the case of unknown singular values we assume that we have additional observations of the operator acting on the singular vectors, i.e. for each $j \in \mathbb{N}$ we have additional observations of $T\varphi_j$ contaminated by an additive error. Due to $T\varphi_j = \sum_{k\in\mathbb{N}} \lambda_k\langle\varphi_j, \varphi_k\rangle_{\mathcal{H}}\psi_k = \lambda_j\psi_j$ and the orthonormality of the singular vectors we obtain

$$\langle T\varphi_j, \psi_j\rangle_{\mathcal{H}} = \lambda_j.$$

That is, we assume that additionally to the observations given by (IGSSM) we have at our disposal noisy observations of $(\lambda_j)_{j\in\mathbb{N}}$.

$$Y_j = \lambda_j\theta_j + \varepsilon\xi_j \qquad \text{with} \qquad \xi_j \overset{\text{iid}}{\sim} \mathcal{N}(0,1), j \in \mathbb{N}, \tag{IGSSM}$$

$$X_j = \lambda_j + \sigma\zeta_j \qquad \text{with} \qquad \zeta_j \overset{\text{iid}}{\sim} \mathcal{N}(0,1), j \in \mathbb{N}. \tag{partially known}$$

This model is called the **indirect Gaussian sequence space model with partially known operator** and noise levels $\varepsilon, \sigma > 0$, since the singular vectors are given, but not the singular values. This model was introduced in Cavalier and Hengartner [2005] and is also considered in Johannes and Schwarz [2013] (estimation), Marteau and Sapatinas [2017a] and Kroll [2019a] (testing). A similar model is considered in Butucea et al. [2008]. For a sequence space model with **fully unknown** operator we refer to Efromovich and Koltchinskii [2001], Marteau [2006] and Hoffmann and Reiss [2008]. In Chapter 1 we consider a further generalization of the partially known (IGSSM), where we allow the noise levels ε and σ to depend on the index j, which is then called a **heterogeneous** model.

Examples

We give some concrete examples of statistical models, where an indirect sequence space model appears naturally.

Example (Inference on derivatives.). (compare Alquier et al. [2011], Section 1.1.6.2.) Let $\mathcal{H} = \mathscr{L}^2[0,1)$ be the Hilbert space of (complex-valued) square integrable functions defined on $[0,1)$ equipped with its usual inner product. Consider the Fourier basis e_j, $j \in \mathbb{Z}$ of $\mathscr{L}^2[0,1)$ with $e_j(x) = \exp(2\pi i j x)$ and for a function $h \in \mathscr{L}^2[0,1)$ its Fourier coefficients $f_j := \langle f, e_j \rangle_{\mathcal{H}}$. Assume we are interested in the β-th derivative of a function f, which can be expressed in the terms of the Fourier coefficients as follows

$$f^{(\beta)} = \sum_{j\in\mathbb{Z}} (2\pi i j)^\beta f_j e_j,$$

i.e. the parameter of interest is given by the Fourier coefficients of the β-th derivative

$$\theta_j = (2\pi i j)^\beta f_j, \qquad j \in \mathbb{Z}$$

but we only have at our disposal the sequence space analogue to a noisy observation of the sequence $f_j = (2\pi i j)^{-\beta}\theta_j$, $j \in \mathbb{N}$. This is a mildly ill-posed problem, where we have complete knowledge of the sequence $(\lambda_j)_{j\in\mathbb{N}} = \left((2\pi i j)^{-\beta}\right)_{j\in\mathbb{N}}$, since it is predetermined by the βth-derivative that we are interested in.

Next, we present an example, where the singular value decomposition of the considered operator is fully known.

Example (Tomography). (compare Baumeister and Leitao [2005], Chapter 5 and Alquier et al. [2011], Section 1.1.6.5. and the references therein) In tomography a goal is to obtain an image of the structure of an object. We assume this structure is characterised by the function f. In X-ray tomography we observe the initial (I_0) and the final (I_1) intensity of an X-ray passing through the object of interest. We denote by $I(x)$ the intensity at a point x, the standard mathematical model for the behaviours of the intensity is given by

$$\frac{I'(x)}{I(x)} = -f(x)$$

for any point x in the object, which implies

$$-\int_{x_0}^{x_1} f(x)\mathrm{d}x = \int_{x_0}^{x_1} \frac{I'(x)}{I(x)}\mathrm{d}x = [\log(I(x))]_{x_0}^{x_1} = \log\left(\frac{I_1}{I_0}\right),$$

where x_0 and x_1 denote the starting and the end point of a line. Hence, observing the initial

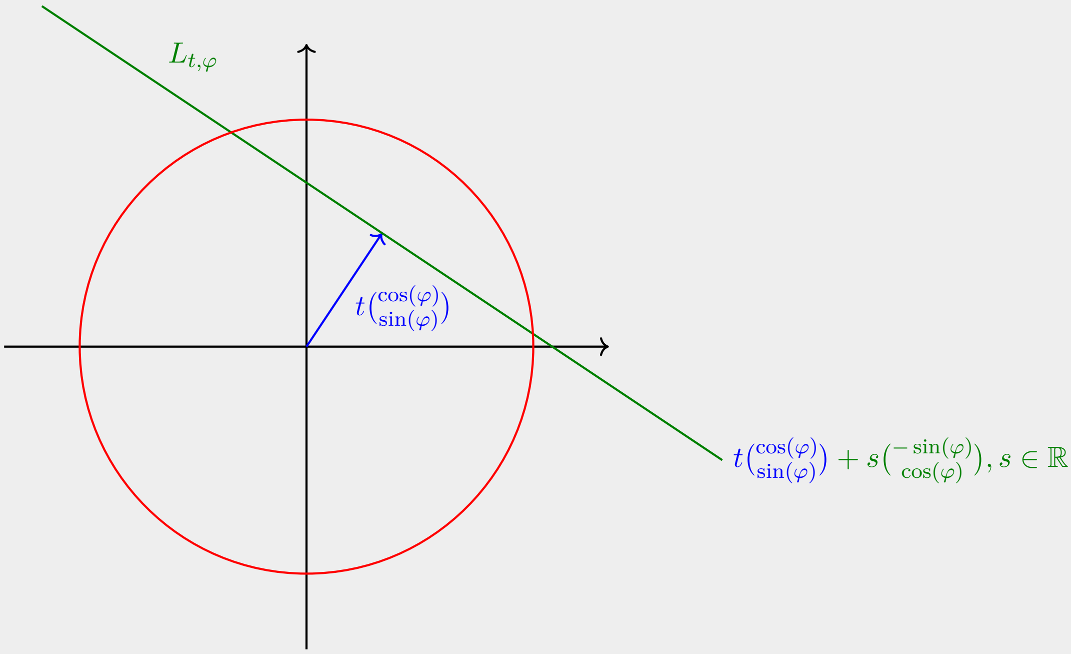
I_0 and final I_1 intensity is equivalent to observing line integrals

$$\exp(-\int_{x_0}^{x_1} f(x)\mathrm{d}x)$$

of the function of interest f. Let us formalize this model by writing it as a linear operator equation. Assume for simplicity that our object is contained in the two-dimensional unit sphere $B := \{x \in \mathbb{R}^2 \mid x_1^2 + x_2^2 \le 1\}$. We consider the following parametrization of a line

$$L_{t,\varphi} = \left\{ z \in \mathbb{R}^2 \mid z = t\begin{pmatrix}\cos\varphi \\ \sin\varphi\end{pmatrix} + s\begin{pmatrix}-\sin\varphi \\ \cos\varphi\end{pmatrix}, s \in \mathbb{R} \right\}$$

for $t \in [0, 1]$ and $\varphi \in [0, 2\pi)$, which is visualized below.



The line integrals of a function $f : B \longrightarrow \mathbb{R}$ are given by

$$\int_{L_{t,\varphi}\cap B} f(z)\mathrm{d}z = \int_{-\sqrt{1-t^2}}^{\sqrt{1-t^2}} f\left(t\begin{pmatrix}\cos\varphi \\ \sin\varphi\end{pmatrix} + s\begin{pmatrix}-\sin\varphi \\ \cos\varphi\end{pmatrix}\right)\mathrm{d}s,$$

and the line integrals can be viewed as functions of (t, φ). The **Radon transform** is defined as the operator that maps a function f to the line integrals, formally

$$\begin{aligned} R : \mathscr{L}^2(B) &\longrightarrow \mathscr{L}^2([0,1] \times [0, 2\pi], \mathrm{d}\mu) \\ f &\longmapsto Rf(\cdot,\cdot) \end{aligned}$$

with

$$Rf(t,\varphi) := \frac{\pi}{2\sqrt{1-t^2}} \int_{L_{t,\varphi}\cap B} f(z)\mathrm{d}z,$$

i.e. $Rf(t, \varphi)$ is the π-the average of f along the line $L_{t,\varphi}$ restricted to the unit circle. By $\mathscr{L}^2([0, 1]\times[0, 2\pi], \mathrm{d}\mu)$ we denote the square integrable functions with respect to the measure $\mu = \frac{2\sqrt{1-t^2}}{\pi}\mathrm{d}t\mathrm{d}s$ (which simply reverses the normalization). It can be shown that the Radon transform R is a linear, bounded and compact operator and its singular value decomposition is known.

The last example presents a case in which the singular value decomposition is only partially given.

Example ((Circular) convolution). (compare Werner [2006], Exercise VI.7.8.) We consider the Hilbert space $\mathcal{H} = \mathscr{L}^2_{\text{per}}[0,1)$ of 1-periodic complex-valued functions defined on $\mathbb{R}$ that are square integrable on $[0,1)$. Let $f \in \mathscr{L}^2_{\text{per}}[0,1)$ be fixed (but unknown) and define the operator

$$T : \mathcal{H} \longrightarrow \mathcal{H}, h \longmapsto Th$$

where

$$Th(x) := \int_{[0,1)} h(t) f(x-t) \mathrm{d}t.$$

It is straight-forward to see that T is a well-defined linear compact operator. Indeed, it is an integral operator with square integrable kernel $k(x,t) = f(x-t)$ (Werner [2006], Example II.3.(c), p.67). Thus, its admits a singular value decomposition. Again, denote by e_j, $j \in \mathbb{Z}$ with $e_j(x) = \exp(2\pi i j x)$, $x \in \mathbb{R}$ the Fourier basis of $\mathscr{L}^2_{\text{per}}[0,1)$ and by $f_j := \langle f, e_j \rangle_{\mathcal{H}}$ the Fourier coefficients of the convolution function f. Then, $(e_j, e_j, f_j)_{j \in \mathbb{Z}}$ is a singular value decomposition of T. It is remarkable that the singular basis $(e_j)_{j \in \mathbb{Z}}$ does not depend on the function f that we convolve with. So even if f is unknown, we have **partial** knowledge about the operator – its singular vectors are known, its singular values are unknown. Depending on the regularity of f, this is either a mildly ill-posed (in the case when f is *ordinary smooth*) or a severely ill-posed model (in the case when f is *super smooth*) with partially known operator.

Additive noise

Let us briefly motivate why the model assumption of additive noise is natural in the statistical context. An extremely well-studied statistical model is nonparametric regression, which can be found in the following simplified form in many textbooks (e.g. Tsybakov [2009]). Based on observations of the equally-spaced point evaluations of a (square-integrable) function $f \in \mathscr{L}^2[0,1)$ defined on $[0,1)$, i.e.

$$Z_j = f\left(\frac{j}{n}\right) + \zeta_j, \qquad \text{where} \qquad \zeta_j \overset{\text{iid}}{\sim} \mathcal{N}(0,1), \qquad \text{for } j \in \{1, \dots, n\}, \tag{NR}$$

we aim to make inference on the function f. The additive noise structure when observing functions is a widely accepted model assumption. We want to explain how this assumption can be transferred to the sequence space model. For any function $\varphi \in \mathscr{L}^2[0,1)$ we define the vectors of point evaluations $\boldsymbol{\varphi} = (\varphi\left(\frac{j}{n}\right))_{j \in \{1,\dots,n\}}$ in $\mathbb{R}^n$. Moreover, we equip the space $\mathbb{R}^n$ with the inner product $\langle \boldsymbol{a}, \boldsymbol{b} \rangle_n = \frac{1}{n} \sum_{j=1}^n a_j b_j$. Let $(\boldsymbol{\varphi_m})_{m \in \mathbb{N}}$ be an orthonormal basis of $\mathscr{L}^2[0,1)$ that satisfies $\langle \boldsymbol{\varphi_l}, \boldsymbol{\varphi_m} \rangle_n = \frac{1}{n} \sum_{j=1}^n \varphi_l\left(\frac{j}{n}\right) \varphi_m\left(\frac{j}{n}\right) = \delta_{l,m}$ for all $l, m \in \{1, \dots, n-1\}$, i.e. its associated vectors are also orthonormal in $(\mathbb{R}^n, \langle \cdot, \cdot \rangle_n)$ (one could e.g. consider the trigonometric basis (cp. Tsybakov [2009], Lemma 1.7.)). We define for $\boldsymbol{Z} = (Z_1, \dots, Z_n)$, $\boldsymbol{\zeta} = (\zeta_1, \dots, \zeta_n)$ and

$k \in \{1, \ldots, n\}$ the quantities

$$Y_k := \langle \boldsymbol{Z}, \boldsymbol{\varphi_k} \rangle_n = \tfrac{1}{n} \sum_{j=1}^{n} Z_j \varphi_k(\tfrac{j}{n}),$$

$$\vartheta_k := \langle \boldsymbol{f}, \boldsymbol{\varphi_k} \rangle_n = \tfrac{1}{n} \sum_{j=1}^{n} f(\tfrac{j}{n}) \varphi_k(\tfrac{j}{n}),$$

$$\xi_k := \sqrt{n} \langle \boldsymbol{\zeta}, \boldsymbol{\varphi_k} \rangle_n = \tfrac{1}{\sqrt{n}} \sum_{j=1}^{n} \zeta_j \varphi_k(\tfrac{j}{n}) \overset{\text{iid}}{\sim} \mathcal{N}(0, 1).$$

Then (NR) implies with $\varepsilon = \frac{1}{\sqrt{n}}$ that

$$Y_k = \vartheta_k + \varepsilon \xi_k, \qquad \text{for } k \in \{1, \ldots, n\}. \tag{TSSM}$$

Note that $\vartheta_k = \langle \boldsymbol{f}, \boldsymbol{\varphi_k} \rangle_n$, $k \in \{1, \ldots, n\}$ are discrete versions of the Fourier coefficients $\theta_k = \int_{[0,1)} f(x) \varphi_k(x) \mathrm{d}x$. Hence, (TSSM) can be viewed as a truncated version of the Gaussian sequence space model, which is well approximated for the idealized case $n \to \infty$. Considering observations as in (NR) of equally spaced point evaluations of Tf in an inverse model, we obtain a similar motivation for the additive error the inverse sequence space model.

Notation

With the index $\bullet$ we indicate sequences. For two sequences $x_\bullet = (x_j)_{j \in \mathbb{N}}$, $y_\bullet = (y_j)_{j \in \mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ operations and inequalities are defined component-wise, i.e. $x_\bullet^2 = (x_j^2)_{j \in \mathbb{N}}$, $x_\bullet / y_\bullet = (x_j / y_j)_{j \in \mathbb{N}}$ and for $c \in \mathbb{R}$, $x_\bullet \leq c y_\bullet$ if $x_j \leq c y_j$ for all $j \in \mathbb{N}$. We denote

$$\ell^2 := \ell^2(\mathbb{N}) := \left\{ x_\bullet \in \mathbb{R}^{\mathbb{N}} : \sum_{j \in \mathbb{N}} x_j^2 < \infty \right\},$$

$$\ell^\infty := \ell^\infty(\mathbb{N}) := \left\{ x_\bullet \in \mathbb{R}^{\mathbb{N}} : \sup_{j \in \mathbb{N}} |x_j| < \infty \right\}.$$

The space $\ell^2 := \ell^2(\mathbb{N})$ equipped with $\langle x_\bullet, y_\bullet \rangle_{\ell^2} := \sum_{j \in \mathbb{N}} x_j y_j$, $\|x_\bullet\|_{\ell^2}^2 := \sum_{j \in \mathbb{N}} x_j^2$ is a Hilbert space of square summable sequences, ℓ^∞ equipped with $\|x_\bullet\|_{\ell^\infty} = \sup_{j \in \mathbb{N}} |x_j|$ is a Banach space of bounded sequences. For $\mathcal{K} \subseteq \mathbb{N}$ we denote the smallest minimiser, if it exists, by

$$\arg\min_{k \in \mathcal{K}} x_k := \min \{ k \in \mathcal{K} : x_k \leq x_j \text{ for all } j \in \mathcal{K} \}.$$

For two functions $f, g : \mathbb{R} \to \mathbb{R}_{>0}$ we write $f(\varepsilon) \lesssim g(\varepsilon)$ (as $\varepsilon \longrightarrow 0$) if there exists a constant $C > 0$ such that $f(\varepsilon) \leq C g(\varepsilon)$ for all ε small enough. We write $f(\varepsilon) \sim g(\varepsilon)$ if both $f(\varepsilon) \lesssim g(\varepsilon)$ and $g(\varepsilon) \lesssim f(\varepsilon)$ and say that f and g are **of the same order**.

Moreover, by $\mathscr{L}^2 := \mathscr{L}^2[0, 1)$ we denote the Hilbert space of real-valued square integrable functions defined on the half-open unit interval $[0, 1)$ equipped with the inner product $\langle f, g \rangle_{\mathscr{L}^2} = \int_0^1 f(x) g(x) \mathrm{d}x$.

Finally, for $k \in \mathbb{N}$ we use the short-hand notation $[\![k]\!]$ for the set $\{1, \ldots, k\}$.

Chapter 1

Adaptive minimax testing with partially known operators

1.1 Introduction

In an inverse Gaussian sequence space model with additional noisy observations of the operator we derive upper bounds for the non-asymptotic minimax radii of testing for ellipsoid-type alternatives, simultaneously for both the signal detection problem (testing against zero) and the goodness-of-fit testing problem (testing against a prescribed sequence) without any regularity assumption on the null hypothesis. The radii are the maximum of two terms, each of which depends on one of the noise levels. Interestingly, the term involving the noise level of the operator explicitly depends on the null hypothesis and vanishes in the signal detection case. We provide a matching lower bound in the case when the operator is observed with the same or smaller noise level as the sequence of interest. We consider two testing procedures, an indirect test based on the estimation of the distance to the null and a direct test, which is instead based on estimating the energy in the image space. We highlight the assumptions under which the direct test performs as well as the indirect test. Furthermore, we apply a classical Bonferroni method for making both the indirect and the direct test adaptive with respect to the regularity of the alternative and derive separation radii for these tests. The radii of the adaptive tests are deteriorated by an additional log-factor, which we show to be unavoidable. The results are illustrated considering Sobolev spaces and mildly or severely ill-posed inverse problems.

The statistical model. We consider an inverse Gaussian sequence space model with heteroscedastic errors and unknown operator

$$
\begin{aligned}
Y_j &= \lambda_j \theta_j + \varepsilon_j \xi_j, \\
X_j &= \lambda_j + \sigma_j \tilde{\xi}_j, \qquad j \in \mathbb{N},
\end{aligned}
\tag{1.1.1}
$$

where $\lambda_\bullet \in \ell^\infty$ is an unknown bounded sequence, $\theta_\bullet \in \ell^2$ is an unknown square summable sequence, $\varepsilon_\bullet \in \mathbb{R}_+^{\mathbb{N}}$ and $\sigma_\bullet \in \mathbb{R}_+^{\mathbb{N}}$ are known sequences of positive real numbers, called noise levels, and $\xi_j, \tilde{\xi}_j \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$. The sequences $Y_\bullet$ and $X_\bullet$ are therefore independent with independent Gaussian components, we denote their respective distributions by $Y_\bullet \sim \mathbb{P}^{\varepsilon_\bullet}_{\lambda_\bullet \theta_\bullet}$ and $X_\bullet \sim \mathbb{P}^{\sigma_\bullet}_{\lambda_\bullet}$ and their joint distribution by $(Y_\bullet, X_\bullet) \sim \mathbb{P}^{\varepsilon_\bullet;\sigma_\bullet}_{\theta_\bullet,\lambda_\bullet}$.

Related literature. Model (1.1.1) is an idealised formulation of a statistical inverse problem with unknown operator, where a signal $\theta_\bullet$ that is transformed by a multiplication with the unknown sequence $\lambda_\bullet$ is observed. In the particular case $\lambda_\bullet = (1)_{j\in\mathbb{N}}$ the model is called direct, otherwise inverse, and ill-posed if additionally $\lambda_\bullet$ tends to zero. The model is already introduced in more detail in the previous section. For statistical inference for inverse problems with fully known operator (corresponding to known $\lambda_\bullet$) we refer to Johnstone and Silverman [1990], Mair and Ruymgaart [1996], Mathé and Pereverzev [2001], Cavalier and Tsybakov [2002], Cavalier et al. [2002] and the references therein. Ingster et al. [2012b] describe examples, in which the inverse Gaussian sequence space model with known $\lambda_\bullet$ arises naturally, one of which is deconvolution (Ermakov [1990], Fan [1991], Stefanski and Carroll [1990]). In our model (1.1.1) the sequence $\lambda_\bullet$ is unknown, but an additional noisy observation of it is available. Cavalier and Hengartner [2005], Ingster et al. [2012a], Johannes and Schwarz [2013] or Marteau and Sapatinas [2017a], for instance, provide a detailed discussion and motivation of this particular statistical inverse problem with unknown operator. An example is density deconvolution with unknown error distribution (c.f. Comte and Lacour [2011], Efromovich [1997] or Neumann [1997]). Oracle or minimax optimal non-parametric estimation and adaptation in the framework of inverse problems has been extensively studied in the literature (see Efromovich and Koltchinskii [2001], Cavalier et al. [2004], Cavalier [2008], Hoffmann and Reiss [2008], to name but a few). In this chapter we are, however, concerned with non-parametric testing, which we formalize next.

The testing task. For some benchmark sequence $\theta^\circ_\bullet \in \ell^2$ we want to test the null hypothesis $\{\theta_\bullet = \theta^\circ_\bullet\}$ against the alternative $\{\theta_\bullet \neq \theta^\circ_\bullet\}$ based on the observations $(Y_\bullet, X_\bullet)$, where $\lambda_\bullet \in \ell^\infty$ is a nuisance parameter. To make the null hypothesis distinguishable from the alternative, we introduce the **separation condition**

$$\ell^2_\rho := \left\{\theta_\bullet \in \ell^2 : \|\theta_\bullet\|^2_{\ell^2} \geq \rho^2\right\},$$

which separates the hypotheses in the ℓ^2-norm by a **separation radius** $\rho > 0$. Additionally, **regularity conditions** are imposed on the unknown sequences $\theta_\bullet$ and $\lambda_\bullet$ by introducing non-parametric classes of parameters $\Theta \subseteq \ell^2$ and $\Lambda \subseteq \ell^\infty$. We define these classes below such that they are flexible enough to capture typical smoothness and ill-posedness assumptions. Over these classes we can write the testing problem as

$$H_0 : \theta_\bullet = \theta^\circ_\bullet, \lambda_\bullet \in \Lambda \qquad \text{against} \qquad H_1^\rho : \theta_\bullet - \theta^\circ_\bullet \in \ell^2_\rho \cap \Theta, \lambda_\bullet \in \Lambda. \tag{1.1.2}$$

Roughly speaking, in minimax testing one searches for the smallest ρ such that (1.1.2) is still testable with small error probabilities. Following e.g. Collier et al. [2017] we measure the accuracy of a test $\Delta : \mathbb{R}^\mathbb{N} \times \mathbb{R}^\mathbb{N} \longrightarrow \{0, 1\}$ by its maximal risk defined as the sum of the maximal type I and II error probability over the null hypothesis and the ρ-separated alternative, respectively,

$$\mathcal{R}(\Delta \mid \Theta, \Lambda, \theta^\circ_\bullet, \rho) := \sup_{\lambda_\bullet \in \Lambda} \mathbb{P}^{\varepsilon_\bullet;\sigma_\bullet}_{\theta^\circ_\bullet, \lambda_\bullet}(\Delta = 1) + \sup_{\substack{\lambda_\bullet \in \Lambda \\ \theta_\bullet - \theta^\circ_\bullet \in \ell^2_\rho \cap \Theta}} \mathbb{P}^{\varepsilon_\bullet;\sigma_\bullet}_{\theta_\bullet, \lambda_\bullet}(\Delta = 0).$$

and compare it to the minimax risk

$$\mathcal{R}(\Theta, \Lambda, \theta^\circ_\bullet, \rho) := \inf_\Delta \mathcal{R}(\Delta \mid \Theta, \Lambda, \theta^\circ_\bullet, \rho),$$

where the infimum is taken over all possible tests, i.e. over all measurable functions $\Delta : \mathbb{R}^\mathbb{N} \times \mathbb{R}^\mathbb{N} \longrightarrow \{0, 1\}$. A separation radius ρ^2 (depending on the classes Θ, Λ, the noise levels $\varepsilon_\bullet, \sigma_\bullet$ and the null hypothesis $\theta^\circ_\bullet$) is called minimax radius of testing, if for all $\alpha \in (0, 1)$ there exist constants $\underline{A}_\alpha \in \mathbb{R}_+$ and $\overline{A}_\alpha \in \mathbb{R}_+$ with

(i) for all $A \geq \overline{A}_\alpha$ we have $\mathcal{R}(\Theta, \Lambda, \theta^\circ_\bullet, A\rho) \leq \alpha$, (upper bound)

(ii) for all $A \leq \underline{A}_\alpha$ we have $\mathcal{R}(\Theta, \Lambda, \theta^\circ_\bullet, A\rho) \geq 1 - \alpha$. (lower bound)

Note that this definition of the minimax radius of testing is entirely non-asymptotic. However, in our illustrations we compare our findings to existing asymptotic results by considering the homoscedastic case, i.e. constant noise levels $\varepsilon_\bullet = (\varepsilon)_{j\in\mathbb{N}}$ and $\sigma_\bullet = (\sigma)_{j\in\mathbb{N}}$ with $\varepsilon, \sigma \in \mathbb{R}_+$, and the behaviour of the radii for ε and σ tending to zero.

Related literature. Minimax testing for the **direct homoscedastic** version of the model (1.1.1), i.e. $\lambda_\bullet = (1)_{j\in\mathbb{N}}, \sigma_\bullet = (0)_{j\in\mathbb{N}}$ and $\varepsilon_\bullet = (\varepsilon)_{j\in\mathbb{N}}$, has been studied extensively in the literature for various classes of alternatives. Asymptotic results and a list of references can be found in the book by Ingster and Suslina [2012]. Let us briefly mention some further references. Lepski and Spokoiny [1999] derive asymptotic minimax rates for Besov-type alternatives. Following this result, Spokoiny [1996] considers adaptive testing strategies, showing that asymptotic adaptation comes with the unavoidable cost of a log-factor. Baraud [2002] introduces a non-asymptotic framework for minimax testing and derives matching upper and lower bounds in the direct model for ellipsoid-type alternatives. Collier et al. [2017] provide similar results for sparse alternatives, using tests based on minimax-optimal estimators of the squared norm of the parameter of interest. Carpentier and Verzelen [2019] derive minimax radii of testing for composite (null) hypotheses, which explicitly depend on the complexity of the null hypothesis. Both phenomena – an estimator of the squared norm yields a minimax optimal test and minimax radii depend on the null hypothesis – reappear in our results.
In the inverse problem setting with **fully known operator** and **homoscedastic** errors, i.e. $\sigma_\bullet = (0)_{j\in\mathbb{N}}$ and $\varepsilon_\bullet = (\varepsilon)_{j\in\mathbb{N}}$ asymptotic rates over ellipsoids Θ are derived in Ingster et al. [2012a]. Simultaneously, Laurent et al. [2012] establish the corresponding non-asymptotic radii. Moreover, Laurent et al. [2011] compare direct and indirect testing approaches, i.e. based on the estimation of $\|\lambda_\bullet(\theta_\bullet - \theta^\circ_\bullet)\|^2_{\ell^2}$ respectively of $\|\theta_\bullet - \theta^\circ_\bullet\|^2_{\ell^2}$, concluding that the direct approach is preferable (under certain assumptions), since it achieves the minimax radius without requiring an inversion. Marteau and Mathé [2014] also discuss how to obtain direct and indirect tests using general regularization schemes. Marteau and Sapatinas [2017b] derive separation radii under weak (non-Gaussian) noise assumptions.
Let us now return to the testing task (1.1.2) in the model with **unknown operator** with **homoscedastic** errors. In this situation there is a natural distinction between the cases $\theta^\circ_\bullet = 0_\bullet := (0)_{j\in\mathbb{N}}$ (*signal detection*) and $\theta^\circ_\bullet \neq 0_\bullet$ (*goodness-of-fit*) on which we comment further below in the next paragraph. Minimax testing in this model is considered in Marteau and Sapatinas [2017a] (only goodness-of-fit) and Kroll [2019a] (goodness-of-fit and signal detection, but treated separately). In the goodness-of-fit scenario Marteau and Sapatinas [2017a] additionally impose an abstract smoothness condition $\theta^\circ_\bullet \in \Theta$ on the null hypothesis and obtain lower and upper bounds featuring a logarithmic gap. Treating the signal detection task and the goodness-of-fit testing task separately, Kroll [2019a] establishes matching upper and lower bounds for the minimax radii of testing uniformly over null hypotheses in Θ. Their radii depend on Θ rather than on the given null hypothesis $\theta^\circ_\bullet$. Let us emphasize that though we are working in a similar setting (with the additional generalization to **heteroscedastic errors**) we instead seek radii for a given $\theta^\circ_\bullet$, which are typically much smaller than the uniform ones obtained by Marteau and Sapatinas [2017a] and Kroll [2019a]. Radii or rates of testing, which depend explicitly on the null hypothesis of the testing problem, are often referred to as **local** rates of testing (c.f. Balakrishnan and Wasserman [2019], Balakrishnan and Wasserman [2018] and Wei and Wainwright [2020]) as opposed to **uniform** rates of testing, which are derived for classes of null hypotheses.

Minimax results. In this paper we derive upper bounds for the non-asymptotic minimax radii of testing in the inverse Gaussian sequence space model simultaneously for both signal detection ($\theta^\circ_\bullet = 0_\bullet$) and goodness-of-fit testing ($\theta^\circ_\bullet \neq 0_\bullet$) without any regularity assumption on the null hypothesis $\theta^\circ_\bullet$. For known operators ($\sigma_\bullet = 0_\bullet$) there is typically no distinction between the goodness-of-fit and the signal detection task. Minimax results for the goodness-of-fit task can be obtained from the signal detection task by simply shifting the observations, i.e. considering the sequence $Y_\bullet - \lambda_\bullet\theta^\circ_\bullet$ instead of $Y_\bullet$. This is obviously no longer possible if $\lambda_\bullet$ is unknown and $\theta^\circ_\bullet \neq 0_\bullet$, which motivates the separate treatment of the two problems in Marteau and Sapatinas [2017a] and Kroll [2019a]. To understand the signal detection problem and the goodness-of-fit testing problem simultaneously we mimic the idea of shifting the observations by reparametrising our statistical model via the mapping $t : \mathbb{R}^{\mathbb{N}} \times \mathbb{R}^{\mathbb{N}} \longrightarrow \mathbb{R}^{\mathbb{N}} \times \mathbb{R}^{\mathbb{N}}, (y_\bullet, x_\bullet) \longmapsto (y_\bullet - \theta^\circ_\bullet x_\bullet, x_\bullet)$, which is bijective and known. The components of $\tilde{Y}_\bullet = Y_\bullet - \theta^\circ_\bullet X_\bullet$ are still independent and follow a normal distribution $\tilde{Y}_j = Y_j - \theta^\circ_j X_j \sim \mathcal{N}\left(\lambda_j(\theta_j - \theta^\circ_j), (\varepsilon^\circ_j)^2\right)$, where $(\varepsilon^\circ_j)^2 := \varepsilon_j^2 + (\theta^\circ_j)^2\sigma_j^2$. This reparametrisation already indicates that with respect to the observations of the operator, the **effective** noise level is $\theta^\circ_\bullet \sigma_\bullet$ instead of the **original** noise level $\sigma_\bullet$. Thereby the dependence of the minimax radii on the null hypothesis is explicit. In particular, this shows that the $\sigma_\bullet$-term in the radius vanishes in the signal detection task ($\theta^\circ_\bullet = 0_\bullet$). Furthermore, for $\sigma_\bullet = 0_\bullet$ we recover the minimax radii for known operators, which consequently do not depend on the null hypothesis $\theta^\circ_\bullet$. Using the reparametrised observations $(\tilde{Y}_\bullet, X_\bullet)$ we propose an **indirect** test based on the estimation of a squared weighted ℓ^2-norm of $\theta^\circ_\bullet - \theta_\bullet$. More precisely, we use an estimator that mimics an inversion of $\lambda_\bullet$ by using the class Λ and aims to estimate the quadratic functional $\mathrm{q}_k^2(\theta^\circ_\bullet - \theta_\bullet) := \sum_{j \in [\![k]\!]} \left(\theta^\circ_j - \theta_j\right)^2$. If k is chosen appropriately, the test attains the minimax radius given by a classical trade-off between the variance of the quadratic functional and a bias^2-term. To avoid the inversion, we investigate a **direct** testing procedure inspired by Laurent et al. [2011] that is based on the estimation of the squared ℓ^2-norm of $\lambda_\bullet(\theta^\circ_\bullet - \theta_\bullet)$. In contrast to inverse problems with known operator, we show that the direct approach is not always preferable if the operator is unknown, but characterise situations in which it is. In particular in signal detection the direct test achieves the minimax radius under very mild assumptions. Moreover, its advantage over the indirect test is that it only implicitly depends on the knowledge of the model's ill-posedness characterised by the class Λ via an optimal choice of the dimension parameter k.

Adaptation. For both testing procedures the optimal choice of the dimension parameter k relies on the knowledge of characteristics of the classes Θ and Λ. A classical procedure to circumvent this problem is to aggregate several tests for various dimension parameters k into a *maximum*-test, which rejects the null hypothesis as soon as one of the tests does. We apply this aggregation to both testing procedures and derive the radii of testing of their corresponding max-tests. Thereby, the indirect max-test is adaptive (i.e. assumption-free) with respect to the smoothness of $\theta_\bullet$ characterised by a family of Θ-alternatives. Comparing its radius to the non-adaptive radius, there is a deterioration, which we express in terms of the number of dimension parameters over which we aggregate. Heuristically, the adaptive radius is obtained by magnifying the error level in the non-adaptive radius by an **adaptive factor** (cp. Spokoiny [1996]). Depending on the complexity of the families of Θ-alternatives, we show that adaptive factors of log log- or even log log log-order are possible. The indirect max-test is still only adaptive with respect to the smoothness of $\theta_\bullet$, but explicitly depends on the model's ill-posedness characterised by Λ. In contrast, the direct max-test is adaptive with respect to both smoothness and ill-posedness. Moreover, we provide a general result (Proposition 1.6.1) which allows to show the unavoidability of adaptive factors for general collections of alternatives. Previously, unavoidability results are only known in specific cases (for instance in Spokoiny [1996] provides such a lower bound for specific Besov-type alternatives in the asymptotic setting). We apply the general lower bound result to specific types of alternatives (consisting of classes of ordinary smooth or super smooth Sobolev-type ellipsoids) and prove that the adaptive factors of log log- and log log log-order, which are attained by our max-tests, are an unavoidable cost to pay for adaptation.

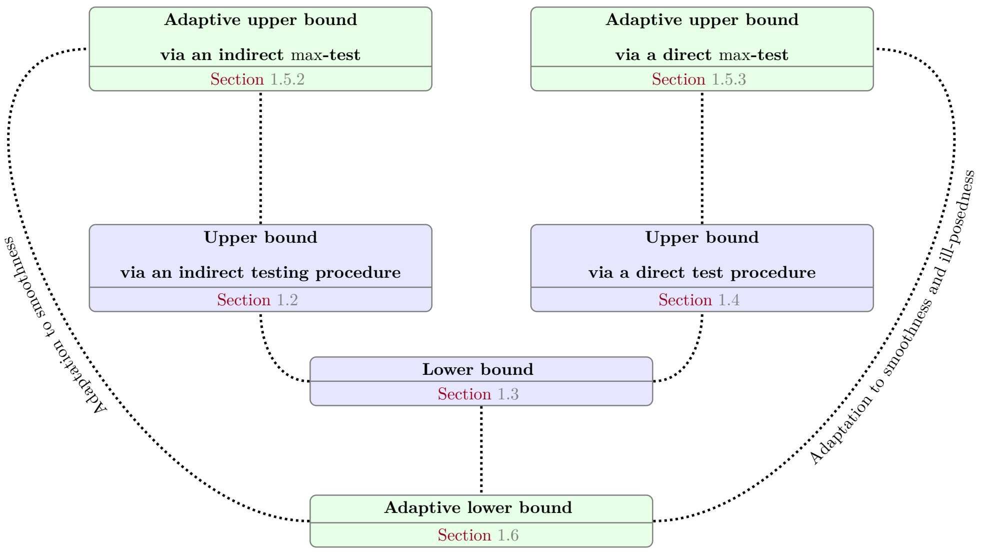
Notation. Due to the many parameters, different noise levels and several radii, the following chapter is heavy on notation. We therefore provide a notation index for easier orientation.

Abbreviations	
$\varepsilon^\circ_\bullet = \sqrt{\varepsilon^2_\bullet + (\theta^\circ_\bullet)^2 \sigma_\bullet}$	reparametrized noise level
$\mathrm{q}^2_k(x_\bullet) = \sum_{j\in[\![k]\!]} x_j^2$	**q**uadratic functional
$\mathrm{m}_k(x_\bullet) := \max_{j\in[\![k]\!]} \lvert x_j \rvert$	**m**aximum (up to k)
$\mathrm{b}_k(\theta_\bullet) := \lVert\theta_\bullet\rVert^2_{\ell^2} - \mathrm{q}^2_k(\theta_\bullet)$	**b**ias terms
$L_u := \sqrt{\lvert\log u\rvert}$	log-term, $u \in (0,1)$
$\delta_{\mathcal{K}} := (1 \vee \log\lvert\mathcal{K}\rvert)^{1/4}$	adaptive factor for aggregation over $\mathcal{K}$
Regularity and ill-posedness	
$a_\bullet \in \mathbb{R}^{\mathbb{N}}_{>0}$	regularity, non-increasing, bounded by 1
$\mathrm{R} > 0$	regularity radius
$\Theta^{\mathrm{R}}_{a_\bullet} \subseteq \ell^2$	regularity class
$\mathrm{v}_\bullet \in \mathbb{R}^{\mathbb{N}}_{>0}$	ill-posedness, non-increasing, bounded by 1
$\mathrm{c} > 0$	ill-posedness diameter
$\Lambda^{\mathrm{c}}_{\mathrm{v}_\bullet} \subseteq \ell^\infty$	ill-posedness class
$\mathcal{A} \subseteq \ell^2$	collection of regularity parameters
$\mathcal{V} \subseteq \ell^\infty$	collection of ill-posedness parameters
Tests, test statistics and thresholds	
$\hat{\mathrm{q}}^2_k$	estimator of $\mathrm{q}^2_k(\theta_\bullet - \theta^\circ_\bullet)$, defined in (1.2.1)
$\tau_k(\alpha)$	**t**hreshold for the indirect test, defined in (1.2.7)
$\Delta_{k,\alpha} := \mathbb{1}_{\{\hat{\mathrm{q}}^2_k > \tau_k(\alpha)\}}$	indirect test, defined in (1.2.8)
$\widetilde{\mathrm{q}}^2_k$	estimator of $\mathrm{q}^2_k(\lambda_\bullet(\theta_\bullet - \theta^\circ_\bullet))$, defined in (1.4.1)
$\tau^{\mathrm{d}}_k(\alpha)$	**t**hreshold for the direct test, defined in (1.4.7)
$\Delta^{\mathrm{d}}_{k,\alpha} = \mathbb{1}_{\{\widetilde{\mathrm{q}}^2_k > \tau^{\mathrm{d}}_k(\alpha)\}}$	direct test, defined in (1.4.8)
$T_{\mathcal{K},\alpha} := \max_{k\in\mathcal{K}} \left\{\hat{\mathrm{q}}^2_k - \tau_k\left(\frac{\alpha}{\lvert\mathcal{K}\rvert}\right)\right\}$	(indirect) max-test statistic (over $\mathcal{K}$)
$\Delta_{\mathcal{K},\alpha} := \mathbb{1}_{\{T_{\mathcal{K},\alpha} > 0\}}$	(indirect) max-test (over $\mathcal{K}$)
$T^{\mathrm{d}}_{\mathcal{K},\alpha} := \max_{k\in\mathcal{K}} \left\{\widetilde{\mathrm{q}}^2_k - \tau^{\mathrm{d}}_k\left(\frac{\alpha}{\lvert\mathcal{K}\rvert}\right)\right\}$	(direct) max-test statistic (over $\mathcal{K}$)
$\Delta^{\mathrm{d}}_{\mathcal{K},\alpha} := \mathbb{1}_{\{T^{\mathrm{d}}_{\mathcal{K},\alpha} > 0\}}$	(direct) max-test (over $\mathcal{K}$)
Separation radii and optimal dimensions	
$\rho^2_{a_\bullet,\mathrm{v}_\bullet}(x_\bullet) := \min_{k\in\mathbb{N}} \left\{\mathrm{q}_k\left(\frac{x^2_\bullet}{\mathrm{v}^2_\bullet}\right) \vee a^2_k\right\}$	(indirect) separation radius
$k_{a_\bullet,\mathrm{v}_\bullet}(x_\bullet) := \arg\min_{k\in\mathbb{N}} \left\{\mathrm{q}_k\left(\frac{x^2_\bullet}{\mathrm{v}^2_\bullet}\right) \vee a^2_k\right\}$	optimal dimension (for the indirect test),
$\left(\rho^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(x_\bullet)\right)^2 := \min_{k\in\mathbb{N}} \left\{\mathrm{v}^{-2}_k \mathrm{q}_k\left(x^2_\bullet\right) \vee a^2_k\right\}$	(direct) separation radius
$k^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(x_\bullet) := \arg\min_{k\in\mathbb{N}} \left\{\mathrm{v}^{-2}_k \mathrm{q}_k\left(x^2_\bullet\right) \vee a^2_k\right\}$	optimal dimension (for the direct test)
$\rho^2_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(x_\bullet) := \min_{k\in\mathcal{K}} \left\{\mathrm{q}_k\left(\frac{x^2_\bullet}{\mathrm{v}^2_\bullet}\right) \vee a^2_k\right\}$	adaptive (indirect) separation radius
$k_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(x_\bullet) := \arg\min_{k\in\mathcal{K}} \left\{\mathrm{q}_k\left(\frac{x^2_\bullet}{\mathrm{v}^2_\bullet}\right) \vee a^2_k\right\}$	optimal dimension contained in $\mathcal{K}$ (indirect)
$r^2_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(x_\bullet) := \min_{k\in\mathcal{K}} \left\{\mathrm{m}_k\left(\frac{x^2_\bullet}{\mathrm{v}^2_\bullet}\right) \vee a^2_k\right\}$	(indirect) remainder radius
$(\rho^{\mathrm{d}}_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(x_\bullet))^2 := \min_{k\in\mathcal{K}} \left\{\mathrm{v}^{-2}_k \mathrm{q}_k(x^2_\bullet) \vee a^2_k\right\}$	adaptive (direct) separation radius
$k^{\mathrm{d}}_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(x_\bullet) := \arg\min_{k\in\mathcal{K}} \left\{\mathrm{v}^{-2}_k \mathrm{q}_k(x^2_\bullet) \vee a^2_k\right\}$	optimal dimension contained in $\mathcal{K}$ (direct)
$(r^{\mathrm{d}}_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(x_\bullet))^2 := \min_{k\in\mathcal{K}} \left\{\mathrm{v}^{-2}_k \mathrm{m}_k(x^2_\bullet) \vee a^2_k\right\}$	(direct) remainder radius

Outline of this chapter. In Section 1.2 we derive an upper bound via an indirect testing procedure, in Section 1.4 we investigate a direct testing procedure. Section 1.3 contains the lower bound. Section 1.5 is devoted to adaptive testing. We describe the general adaptation procedure in Section 1.5.1 and apply it to both the indirect test (Section 1.5.2) and the direct test (Section 1.5.2). An adaptive lower bound can be found in Section 1.6. Technical results and their proofs are deferred to Appendix A.

Outline

Adaptive minimax testing with partially known operators



1.2 Upper bound via an indirect testing procedure

Regularity classes. The tests we propose are based on estimators of quadratic functionals, for any sequence $x_\bullet \in \mathbb{R}^{\mathbb{N}}$ and $k \in \mathbb{N}$ we define

$$\mathrm{q}_k^2(x_\bullet) := \sum_{j\in[\![k]\!]} x_j^2, \qquad \mathrm{m}_k(x_\bullet) := \max_{j\in[\![k]\!]} |x_j| \,.$$

Moreover, for a sequence $\theta_\bullet \in \ell^2$ and $k \in \mathbb{N}$ we define the bias terms

$$\mathrm{b}_k(\theta_\bullet) := \|\theta_\bullet\|_{\ell^2}^2 - \mathrm{q}_k^2(\theta_\bullet) = \sum_{j>k} \theta_j^2.$$

With this notation we are ready to define the non-parametric classes for the parameters $\theta_\bullet \in \ell^2$ and $\lambda_\bullet \in \ell^\infty$. Let $a_\bullet = (a_j)_{j\in\mathbb{N}} \subseteq \mathbb{R}_+^{\mathbb{N}}$ be a strictly positive monotonically non-increasing sequence bounded by 1 and let $\mathrm{R} > 0$, we define

$$\Theta_{a_\bullet}^{\mathrm{R}} := \left\{\theta_\bullet \in \ell^2 : \mathrm{b}_\bullet(\theta_\bullet) \leq \mathrm{R}^2 a_\bullet^2\right\} = \left\{\theta_\bullet \in \ell^2 : \mathrm{b}_k(\theta_\bullet) \leq \mathrm{R}^2 a_k^2 \text{ for all } k \in \mathbb{N}\right\}.$$

Note that $\Theta_{a_\bullet}^{\mathrm{R}}$ is of a very general form, it simply allows to control the bias terms $\mathrm{b}_k, k \in \mathbb{N}$ for all elements in the class, which is sufficient for all our proofs. A more common class of alternatives (e.g. used in Kroll [2019a], Marteau and Sapatinas [2017a], Baraud [2002]) are ellipsoids of the

form $\tilde{\Theta}^{\mathrm{R}}_{a_\bullet} := \left\{\theta_\bullet \in \ell^2 : \sum_{j\in\mathbb{N}} \theta_j^2 a_j^{-2} \le \mathrm{R}^2\right\}$, which is also covered by our class $\Theta^{\mathrm{R}}_{a_\bullet}$. We refer to Tsybakov [2009] (Lemma A.3. in the appendix) for an explanation how the decay of the sequence $\theta_\bullet$, which can be interpreted as coefficients of a function w.r.t. a certain basis, relates to the regularity of the associated function.
Let $\mathrm{v}_\bullet = (\mathrm{v}_j)_{j\in\mathbb{N}} \subseteq \mathbb{R}^{\mathbb{N}}_+$ be a strictly positive, monotonically non-increasing sequence bounded by 1 and let $\mathrm{c} \ge 1$, we define

$$\Lambda^{\mathrm{c}}_{\mathrm{v}_\bullet} := \left\{\lambda_\bullet \in \ell^\infty : \mathrm{c}^{-1} \le \frac{\lambda_\bullet^2}{\mathrm{v}_\bullet^2} \le \mathrm{c}\right\} = \left\{\lambda_\bullet \in \ell^\infty : \mathrm{c}^{-1} \le \frac{\lambda_k^2}{\mathrm{v}_k^2} \le \mathrm{c} \text{ for all } k \in \mathbb{N}\right\}.$$

Let us emphasise that the assumptions $\lambda_\bullet \in \Lambda^{\mathrm{c}}_{\mathrm{v}_\bullet}$ and $\mathrm{v}_\bullet \in \mathbb{R}^{\mathbb{N}}_{>0}$ imply that $\lambda_\bullet > 0_\bullet$ and hence the parameter $\theta_\bullet$ is identifiable.

Definition of the test statistic. In this section we derive an upper bound for the minimax radius of testing based on the estimation of the energy of the parameter of interest $\theta_\bullet - \theta^\circ_\bullet$. To be more precise, for the reparametrised $(\varepsilon^\circ_\bullet)^2 = \varepsilon^2_\bullet + (\theta^\circ_\bullet)^2\sigma^2_\bullet \in \mathbb{R}^{\mathbb{N}}_+$ we consider the estimators

$$\hat{\mathrm{q}}_k^2 := \sum_{j\in[\![k]\!]} \frac{(Y_j - \theta^\circ_j X_j)^2 - (\varepsilon^\circ_j)^2}{\mathrm{v}_j^2} = \sum_{j\in[\![k]\!]} \frac{\tilde{Y}_j^2 - (\varepsilon^\circ_j)^2}{\mathrm{v}_j^2}. \tag{1.2.1}$$

Since

$$\begin{aligned}
\mathbb{E}^{\varepsilon_\bullet,\sigma_\bullet}_{\theta_\bullet,\lambda_\bullet}\left(\hat{\mathrm{q}}_k^2\right) &= \sum_{j\in[\![k]\!]} \frac{\mathbb{E}^{\varepsilon_\bullet,\sigma_\bullet}_{\theta_\bullet,\lambda_\bullet}(Y_j - \theta^\circ_j X_j)^2 - (\varepsilon^\circ_j)^2}{\mathrm{v}_j^2} = \sum_{j\in[\![k]\!]} \frac{(\varepsilon^\circ_j)^2 + \lambda_j^2\left(\theta_j - \theta^\circ_j\right)^2 - (\varepsilon^\circ_j)^2}{\mathrm{v}_j^2} \\
&= \sum_{j\in[\![k]\!]} \frac{\lambda_j^2\left(\theta_j - \theta^\circ_j\right)^2}{\mathrm{v}_j^2} \\
&= \mathrm{q}_k^2\left(\frac{\lambda_\bullet}{\mathrm{v}_\bullet}(\theta_\bullet - \theta^\circ_\bullet)\right) \le \mathrm{c}\mathrm{q}_k^2(\theta_\bullet - \theta^\circ_\bullet),
\end{aligned}$$

$\hat{\mathrm{q}}_k^2$ is an unbiased estimator of the quadratic functional $\mathrm{q}_k^2\left(\frac{\lambda_\bullet}{\mathrm{v}_\bullet}(\theta_\bullet - \theta^\circ_\bullet)\right)$, which differs from $\mathrm{q}_k^2(\theta_\bullet - \theta^\circ_\bullet)$ only by a factor c for all $\lambda_\bullet \in \Lambda^{\mathrm{c}}_{\mathrm{v}_\bullet}$ and all $k \in \mathbb{N}$. For a sequence $x_\bullet \in \mathbb{R}^{\mathbb{N}}$ let us define the following minimum and minimiser, respectively,

$$\rho^2_{a_\bullet,\mathrm{v}_\bullet}(x_\bullet) := \min_{k\in\mathbb{N}}\left\{\mathrm{q}_k\left(\frac{x_\bullet^2}{\mathrm{v}_\bullet^2}\right) \vee a_k^2\right\} = \min_{k\in\mathbb{N}}\left\{\sqrt{\sum_{j\in[\![k]\!]} \frac{x_j^4}{\mathrm{v}_j^4}} \vee a_k^2\right\}, \tag{1.2.2}$$

$$k_{a_\bullet,\mathrm{v}_\bullet}(x_\bullet) := \arg\min_{k\in\mathbb{N}}\left\{\mathrm{q}_k\left(\frac{x_\bullet^2}{\mathrm{v}_\bullet^2}\right) \vee a_k^2\right\} = \arg\min_{k\in\mathbb{N}}\left\{\sqrt{\sum_{j\in[\![k]\!]} \frac{x_j^4}{\mathrm{v}_j^4}} \vee a_k^2\right\}. \tag{1.2.3}$$

Throughout this section the sequences $a_\bullet$ and $\mathrm{v}_\bullet$ are arbitrary but fixed. In particular, the optimal testing procedures explicitly exploit the prior knowledge of $a_\bullet$ and $\mathrm{v}_\bullet$, i.e. the fact that the unknown parameters satisfy $\theta_\bullet - \theta^\circ_\bullet \in \Theta^{\mathrm{R}}_{a_\bullet}$ and $\lambda \in \Lambda^{\mathrm{c}}_{\mathrm{v}_\bullet}$ for some $\mathrm{R}, \mathrm{c} \in \mathbb{R}_+$. Given subsets $\mathcal{A}, \mathcal{V} \subseteq \mathbb{R}^{\mathbb{N}}_+$ of strictly positive, monotonically non-increasing bounded sequences, we discuss adaptive testing strategies when $a_\bullet \in \mathcal{A}$ and $\mathrm{v}_\bullet \in \mathcal{V}$ in Section 1.5.

Our evaluation of the performance of the test under both the null hypothesis and the alternative relies on bounds for quantiles of (non-)central χ^2-distributions, which we present in Lemma A.1.1 in Section A.1. Its proof is based on a result in Birgé [2001] (Lemma 8.1), which is a generalisation of Lemma 1 of Laurent and Massart [2000] and can also be found with slightly different notation in Laurent et al. [2012] (Lemma 2).

Proposition 1.2.1 (Bounds for the quantiles of $\hat{q}_k^2$). For $u \in (0,1)$ we define $L_u := \sqrt{|\log u|}$. Let $\alpha, \beta \in (0,1)$.

(i) **(level-α)** For each $k \in \mathbb{N}$ we have

$$\sup_{\lambda \in \Lambda^{c}_{v_\bullet}} \mathbb{P}^{\varepsilon_\bullet, \sigma_\bullet}_{\theta^\circ_\bullet, \lambda_\bullet}\left(\hat{q}_k^2 > 2L_\alpha q_k^2 \left(\frac{(\varepsilon^\circ_\bullet)^2}{v_\bullet^2} \right) + 2L_\alpha^2 m_k \left(\frac{(\varepsilon^\circ_\bullet)^2}{v_\bullet^2} \right) \right) \leq \alpha. \tag{1.2.4}$$

(ii) **($(1-\beta)$-powerful)** Define the dimension $k_\star := k_{a_\bullet, v_\bullet}(\varepsilon_\bullet) \wedge k_{a_\bullet, v_\bullet}(\theta^\circ_\bullet \sigma_\bullet) \in \mathbb{N}$ as in (1.2.3) and $C_{\alpha,\beta} := 5\left(L_\alpha + L_\alpha^2 + L_\beta + 5L_\beta^2\right)$. Then for each $\theta_\bullet - \theta^\circ_\bullet \in \Theta^{R}_{a_\bullet} \cap \ell^2_\rho$ with

$$\rho^2 \geq \left(R^2 + cC_{\alpha,\beta}\right)\left\{\rho^2_{a_\bullet, v_\bullet}(\varepsilon_\bullet) \vee \rho^2_{a_\bullet, v_\bullet}(\theta^\circ_\bullet \sigma_\bullet)\right\}$$

we have

$$\sup_{\lambda \in \Lambda^{c}_{v_\bullet}} \mathbb{P}^{\varepsilon_\bullet, \sigma_\bullet}_{\theta_\bullet, \lambda_\bullet}\left(\hat{q}_{k_\star}^2 \leq 2L_\alpha q_{k_\star}^2 \left(\frac{(\varepsilon^\circ_\bullet)^2}{v_\bullet^2} \right) + 2L_\alpha^2 m_{k_\star} \left(\frac{(\varepsilon^\circ_\bullet)^2}{v_\bullet^2} \right) \right) \leq \beta. \tag{1.2.5}$$

Proof of Proposition 1.2.1. We intend to apply Lemma A.1.1 and use the notation introduced there. If $(Y_\bullet, X_\bullet) \sim \mathbb{P}^{\varepsilon_\bullet, \sigma_\bullet}_{\theta_\bullet, \lambda_\bullet}$, then for each $k \in \mathbb{N}$,

$$Q_k := \hat{q}_k^2 + q_k^2\left(\frac{\varepsilon^\circ_\bullet}{v_\bullet}\right) = \sum_{j \in [\![k]\!]} \frac{(Y_j - \theta^\circ_j X_j)^2}{v_j^2} \sim \mathbb{Q}^{e_\bullet}_{\mu_\bullet, k}$$

with $e_\bullet := \frac{\varepsilon^\circ_\bullet}{v_\bullet}$ and $\mu_\bullet := \frac{\lambda_\bullet(\theta_\bullet - \theta^\circ_\bullet)}{v_\bullet}$.

(i) Under the null hypothesis $\theta_\bullet = \theta^\circ_\bullet$, i.e., $(Y_\bullet, X_\bullet) \sim \mathbb{P}^{\varepsilon_\bullet, \sigma_\bullet}_{\theta^\circ_\bullet, \lambda_\bullet}$ we have $Q_k \sim \mathbb{Q}^{e_\bullet}_{0_\bullet, k}$. Therefore, with (A.1.1) from Lemma A.1.1 it follows

$$q^{e_\bullet}_{0_\bullet, k}(u) \leq q_k^2(e_\bullet) + 2L_u q_k(e_\bullet^2) + 2L_u^2 m_k(e_\bullet^2),$$

which implies (1.2.4).

(ii) Under the alternative, i.e. $(Y_\bullet, X_\bullet) \sim \mathbb{P}^{\varepsilon_\bullet, \sigma_\bullet}_{\theta_\bullet, \lambda_\bullet}$ with $\lambda_\bullet \in \Lambda^{c}_{v_\bullet}$, $\theta_\bullet - \theta^\circ_\bullet \in \Theta^{R}_{a_\bullet} \cap \ell^2_\rho$ and $\rho^2 \geq (R^2 + cC_{\alpha,\beta})\left\{\rho^2_{a_\bullet, v_\bullet}(\varepsilon_\bullet) \vee \rho^2_{a_\bullet, v_\bullet}(\theta^\circ_\bullet \sigma_\bullet)\right\}$, we obtain

$$\begin{aligned}
\|\theta_\bullet - \theta^\circ_\bullet\|^2_{\ell^2} &\geq \rho^2 \geq \left(R^2 + cC_{\alpha,\beta}\right)\left\{\rho^2_{a_\bullet, v_\bullet}(\varepsilon_\bullet) \vee \rho^2_{a_\bullet, v_\bullet}(\theta^\circ_\bullet \sigma_\bullet)\right\} \\
&\geq R^2 a_{k_\star}^2 + cC_{\alpha,\beta}\left\{ q_{k_\star}\left(\frac{\varepsilon_\bullet^2}{v_\bullet^2}\right) \vee q_{k_\star}\left(\frac{(\theta^\circ_\bullet)^2 \sigma_\bullet^2}{v_\bullet^2}\right)\right\} \\
&\geq R^2 a_{k_\star}^2 + c\frac{C_{\alpha,\beta}}{2}\left\{ q_{k_\star}\left(e_\bullet^2\right)\right\} \\
&\geq R^2 a_{k_\star}^2 + c\frac{5}{2}\left\{ L_\alpha q_{k_\star}\left(e_\bullet^2\right) + L_\alpha^2 m_{k_\star}\left(e_\bullet^2\right) + q_{k_\star}\left(e_\bullet^2\right)(L_\beta + 5L_\beta^2)\right\},
\end{aligned} \tag{1.2.6}$$

using that $\rho^2_{a_\bullet, v_\bullet}(\varepsilon_\bullet) \vee \rho^2_{a_\bullet, v_\bullet}(\theta^\circ_\bullet \sigma_\bullet) = q_{k_\star}\left(\frac{\varepsilon_\bullet^2}{v_\bullet^2}\right) \vee q_{k_\star}\left(\frac{(\theta^\circ_\bullet)^2\sigma_\bullet^2}{v_\bullet^2}\right) \vee a_{k_\star}^2$, which follows from Lemma A.2.1 and $2\left\{q_{k_\star}\left(\frac{\varepsilon_\bullet^2}{v_\bullet^2}\right) \vee q_{k_\star}\left(\frac{(\theta^\circ_\bullet)^2\sigma_\bullet^2}{v_\bullet^2}\right)\right\} \geq q_{k_\star}\left(\frac{(\varepsilon^\circ_\bullet)^2}{v_\bullet^2}\right) = q_{k_\star}(e_\bullet^2) \geq m_{k_\star}(e_\bullet^2)$. Moreover, for each $k \in \mathbb{N}$ and $\lambda \in \Lambda^{c}_{v_\bullet}$ we have $cq_k^2(\mu_\bullet) \geq q_k^2(\theta_\bullet - \theta^\circ_\bullet) = \|\theta_\bullet - \theta^\circ_\bullet\|^2_{\ell^2} - b_k^2(\theta_\bullet - \theta^\circ_\bullet)$, which in turn for each $\theta_\bullet - \theta^\circ_\bullet \in \Theta^{R}_{a_\bullet}$ implies

$$cq_k^2(\mu_\bullet) \geq \|\theta_\bullet - \theta^\circ_\bullet\|^2_{\ell^2} - R^2 a_k^2.$$

This bound applied with $k = k_\star$ together with (1.2.6) yields

$$\mathrm{c}\mathrm{q}_{k_\star}^2(\mu_\bullet) \geq \mathrm{c}\frac{5}{2}\left\{L_\alpha \mathrm{q}_{k_\star}\left(e_\bullet^2\right) + L_\alpha^2 \mathrm{m}_{k_\star}\left(e_\bullet^2\right) + \mathrm{q}_{k_\star}\left(e_\bullet^2\right)(L_\beta + 5L_\beta^2)\right\}.$$

Rearranging the last inequality we obtain

$$\frac{4}{5}\mathrm{q}_{k_\star}^2(\mu_\bullet) \geq 2L_\alpha \mathrm{q}_{k_\star}\left(e_\bullet^2\right) + 2L_\alpha^2 \mathrm{m}_{k_\star}\left(e_\bullet^2\right) + \mathrm{q}_{k_\star}\left(e_\bullet^2\right) 2(L_\beta + 5L_\beta^2).$$

Inserting this bound into (A.1.3) of Lemma A.1.1 implies

$$\begin{aligned} 2L_\alpha \mathrm{q}_{k_\star}\left(e_\bullet^2\right) + 2L_\alpha^2 \mathrm{m}_{k_\star}\left(e_\bullet^2\right) + \mathrm{q}_{k_\star}^2(e_\bullet) &\leq \frac{4}{5}\mathrm{q}_{k_\star}^2(\mu_\bullet) - \mathrm{q}_{k_\star}\left(e_\bullet^2\right) 2(L_\beta + 5L_\beta^2) + \mathrm{q}_{k_\star}^2(e_\bullet) \\ &\leq \mathrm{q}_{\mu_\bullet,k}^{e_\bullet}(1-\beta), \end{aligned}$$

and thus (1.2.5), which completes the proof.

□

Definition of the test. For $\alpha \in (0,1)$ and $k \in \mathbb{N}$ we define the threshold

$$\tau_k(\alpha) := 2L_\alpha \mathrm{q}_k\left(\frac{(\varepsilon_\bullet^\circ)^2}{\mathrm{v}_\bullet^2}\right) + 2L_\alpha^2 \mathrm{m}_k\left(\frac{(\varepsilon_\bullet^\circ)^2}{\mathrm{v}_\bullet^2}\right) \tag{1.2.7}$$

and the corresponding test

$$\Delta_{k,\alpha} := \mathbb{1}_{\{\hat{\mathrm{q}}_k^2 > \tau_k(\alpha)\}}. \tag{1.2.8}$$

Proposition 1.2.1 (i) shows that the test $\Delta_{k,\alpha/2}$ is a level $\alpha/2$-test for any $k \in \mathbb{N}$. Moreover, $\Delta_{k_\star,\alpha/2}$ with

$$k_\star := k_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet) \wedge k_{a_\bullet,\mathrm{v}_\bullet}(\theta_\bullet^\circ \sigma_\bullet) \tag{1.2.9}$$

is a $(1-\alpha/2)$-powerful test over $\overline{A}_\alpha\{\rho_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet) \vee \rho_{a_\bullet,\mathrm{v}_\bullet}(\theta_\bullet^\circ\sigma_\bullet)\}$-separated alternatives due to Proposition 1.2.1 (ii) with $\beta = \alpha/2$ and $\overline{A}_\alpha^2 := \mathrm{R}^2 + \mathrm{c}C_{\alpha/2,\alpha/2} = \mathrm{R}^2 + \mathrm{c}\left(10L_{\alpha/2} + 30L_{\alpha/2}^2\right)$. Hence,

$$\mathcal{R}\left(\Delta_{k_\star,\alpha/2} \mid \Theta_{a_\bullet}^{\mathrm{R}}, \Lambda_{\mathrm{v}_\bullet}^{\mathrm{c}}, \theta_\bullet^\circ, A\{\rho_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet) \vee \rho_{a_\bullet,\mathrm{v}_\bullet}(\theta_\bullet^\circ\sigma_\bullet)\}\right) \leq \alpha/2 + \alpha/2 = \alpha$$

for all $A \geq \overline{A}_\alpha$. In other words, $\rho_{a_\bullet,\mathrm{v}_\bullet}^2(\varepsilon_\bullet) \vee \rho_{a_\bullet,\mathrm{v}_\bullet}^2(\theta_\bullet^\circ\sigma_\bullet)$ is an upper bound for the radius of testing of $\Delta_{k_\star,\alpha/2}$, which is summarised in the next theorem.

Theorem 1.2.2 (Upper bound for the radius of testing). For $\alpha \in (0,1)$ define $\overline{A}_\alpha^2 := \mathrm{R}^2 + \mathrm{c}\left(10L_{\alpha/2} + 30L_{\alpha/2}^2\right)$. Then, for all $A \geq \overline{A}_\alpha$ we have

$$\mathcal{R}\left(\Theta_{a_\bullet}^{\mathrm{R}}, \Lambda_{\mathrm{v}_\bullet}^{\mathrm{c}}, \theta_\bullet^\circ, A\rho\right) \leq \alpha,$$

with $\rho := \rho_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet) \vee \rho_{a_\bullet,\mathrm{v}_\bullet}(\theta_\bullet^\circ\sigma_\bullet)$, i.e. ρ^2 is an upper bound for the minimax radius of testing.

Proof of Theorem 1.2.2. The claim follows from Proposition 1.2.1 considering $\Delta_{k_\star,\alpha/2}$ defined in (1.2.8) and the elementary bound

$$\mathcal{R}\left(\Theta_{a_\bullet}^{\mathrm{R}}, \Lambda_{\mathrm{v}_\bullet}^{\mathrm{c}}, \theta_\bullet^\circ, A\rho\right) \leq \mathcal{R}\left(\Delta_{k_\star,\alpha/2} \mid \Theta_{a_\bullet}^{\mathrm{R}}, \Lambda_{\mathrm{v}_\bullet}^{\mathrm{c}}, \theta_\bullet^\circ, A\rho\right).$$

□

Remark 1.2.3 (Signal detection vs. goodness-of-fit). *Considering the* ***signal detection task****, i.e.* $\theta^\circ_\bullet = 0_\bullet$*, we have* $\rho_{a_\bullet,\mathrm{v}_\bullet}(\theta^\circ_\bullet\sigma_\bullet) = 0$ *for all* $\sigma_\bullet \in \mathbb{R}^{\mathbb{N}}_+$ *and, thus, the minimax radius of testing does not depend on the noise level* $\sigma_\bullet$*.*
Considering the ***goodness-of-fit task****, i.e.* $\theta^\circ_\bullet \neq 0_\bullet$*, for all* $\varepsilon_\bullet \geq \sigma_\bullet$*, we have*

$$\mathrm{q}_k\left(\frac{(\theta^\circ_\bullet)^2\sigma^2_\bullet}{\mathrm{v}^2_\bullet}\right) \leq \mathrm{q}_k\left(\frac{(\theta^\circ_\bullet)^2\varepsilon^2_\bullet}{\mathrm{v}^2_\bullet}\right) \leq \|\theta^\circ_\bullet\|^2_{\ell^\infty}\,\mathrm{q}_k\left(\frac{\varepsilon^2_\bullet}{\mathrm{v}^2_\bullet}\right)$$

and, therefore, $\rho_{a_\bullet,\mathrm{v}_\bullet}(\theta^\circ_\bullet\sigma_\bullet) \leq \|\theta^\circ_\bullet\|_{\ell^\infty}\,\rho_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet)$*. In other words,* $\rho_{a_\bullet,\mathrm{v}_\bullet}(\theta^\circ_\bullet\sigma_\bullet)$ *is negligible compared to* $\|\theta^\circ_\bullet\|_{\ell^\infty}\,\rho_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet)$*. We point out that* $\varepsilon_\bullet \geq \sigma_\bullet$ *is a natural situation. It essentially means that the (multiplication) operator can be observed with the same or at lower noise level than the sequence of interest. A similar assumption is for instance also considered in Cavalier and Hengartner [2005] and Efromovich [1997]. Often, if both measurements* $Y_\bullet$ *and* $X_\bullet$ *are made with the same "measurement device", it is even common to assume* $\varepsilon_\bullet = \sigma_\bullet$*.* □

Remark 1.2.4 (Homoscedastic, (non-)parametric rates). *In the homoscedastic case, i.e.* $\varepsilon_\bullet = (\varepsilon)_{j\in\mathbb{N}}$ *and* $\sigma_\bullet = (\sigma)_{j\in\mathbb{N}}$ *for* $\varepsilon, \sigma \in \mathbb{R}_+$ *we are especially interested in the behaviour of the radii of testing* $\rho_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon) := \rho_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon^\circ_\bullet)$ *and* $\rho_{a_\bullet,\mathrm{v}_\bullet}(\sigma) := \rho_{a_\bullet,\mathrm{v}_\bullet}(\theta^\circ_\bullet\sigma_\bullet)$ *as* ε *and* σ *tend to zero.* $\rho_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon)$ *and* $\rho_{a_\bullet,\mathrm{v}_\bullet}(\sigma)$ *are then called* ***rates of testing****. We call* $\rho_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon)$ *(respectively* $\rho_{a_\bullet,\mathrm{v}_\bullet}(\sigma)$*) parametric, if* $\frac{\rho_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon)}{\varepsilon}$ *is bounded away from* 0 *and infinity as* $\varepsilon \to 0$*. Note that since*

$$\liminf_{\varepsilon\to 0}\frac{\rho_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon)}{\varepsilon} \geq \|\mathrm{v}_\bullet\|^{-2}_\infty$$

and $a_\bullet > 0_\bullet$*, it is always bounded away from* 0*. Hence, it becomes parametric if and only if* $\mathrm{v}^{-2}_\bullet \in \ell^2$*. However, since* $\mathrm{v}_\bullet \in \ell^\infty$ *and therefore* $\mathrm{v}^{-2}_j \geq \|\mathrm{v}_\bullet\|^{-2}_\infty$ *for all* $j \in \mathbb{N}$*, we always have* $\mathrm{v}^{-2}_\bullet \notin \ell^2$*. Thus, the rate* $\rho_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon)$ *is always* ***non-parametric.***
On the other hand, for a goodness-of-fit task ($\theta^\circ_\bullet \neq 0_\bullet$*), it can similarly be seen that the rate* $\rho_{a_\bullet,\mathrm{v}_\bullet}(\sigma)$ *is parametric if and only if* $\frac{\theta^\circ_\bullet}{\mathrm{v}_\bullet} \in \ell^2$*, which is possible. Note that it is never faster than parametric, since*

$$\liminf_{\sigma\to 0}\frac{\rho_{a_\bullet,\mathrm{v}_\bullet}(\sigma)}{\sigma} \geq \left\|(\theta^\circ_\bullet)^2\right\|_{\ell^2}\|\mathrm{v}_\bullet\|^{-2}_\infty > 0.$$

□

Illustration 1.2.5. Throughout this chapter we illustrate the order of the rates of testing in the homoscedastic case $\varepsilon_\bullet = (\varepsilon)_{j\in\mathbb{N}}$ and $\sigma_\bullet = (\sigma)_{j\in\mathbb{N}}$ under the following typical smoothness and ill-posedness assumptions. Concerning the regularity class $\Theta^{\mathrm{R}}_{a_\bullet}$ we distinguish two behaviours of the sequence $a_\bullet$, namely the **ordinary smooth** case $a_\bullet = (j^{-s})_{j\in\mathbb{N}}$ for $s > 1/2$, where $\Theta^{\mathrm{R}}_{a_\bullet}$ corresponds to a Sobolev ellipsoid, and the **super smooth** case $a_\bullet = (e^{-j^s})_{j\in\mathbb{N}}$ for $s > 0$, which can be interpreted as an analytic class of parameters. Concerning the class $\Lambda^{\mathrm{c}}_{\mathrm{v}_\bullet}$ we also distinguish two cases for the sequence $\mathrm{v}_\bullet$. For $p > 0$ we consider a **mildly ill-posed** model $\mathrm{v}_\bullet = (j^{-p})_{j\in\mathbb{N}}$ and a severely ill-posed model $\mathrm{v}_\bullet = (e^{-j^p})_{j\in\mathbb{N}}$. Finally, we consider two cases of null hypotheses: the **signal detection task** $\theta^\circ_\bullet = 0_\bullet$ and the **goodness-of-fit testing task** $\theta^\circ_\bullet = (j^{-t})_{j\in\mathbb{N}}$ for some $t > 1/2$. The table below displays the order of the optimal choice $k_\star := k_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet) \wedge k_{a_\bullet,\mathrm{v}_\bullet}(\theta^\circ_\bullet\sigma_\bullet)$ for the dimension parameter as well as the order of the minimax rate $\rho^2_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet) \vee \rho^2_{a_\bullet,\mathrm{v}_\bullet}(\theta^\circ_\bullet\sigma_\bullet)$ for the signal detection task (with $\rho^2_{a_\bullet,\mathrm{v}_\bullet}(\theta^\circ_\bullet\sigma_\bullet) = 0$ as discussed in Remark 1.2.3) and the goodness-of-fit task. Keep in mind that the rate $\rho^2_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet)$ does not depend on the null hypothesis, therefore, it is the same for all $\theta^\circ_\bullet \in \ell^2$. In accordance with Remark 1.2.4, $\rho^2_{a_\bullet,\mathrm{v}_\bullet}(\theta^\circ_\bullet\sigma_\bullet)$ is parametric for the goodness-of-fit task whenever $\frac{(\theta^\circ_\bullet)^2}{\mathrm{v}^2_\bullet} \in \ell^2$. The calculations of the order of the radii in this chapter are similar to the calculations in the illustrations in Chapter 3 and Chapter 4 (replace n with ε^2) and thus omitted.

Order of the optimal dimension $k_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet) \wedge k_{a_\bullet,\mathrm{v}_\bullet}(\theta^\circ_\bullet\sigma_\bullet)$ and the minimax radius $\rho^2_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet) \vee \rho^2_{a_\bullet,\mathrm{v}_\bullet}(\theta^\circ_\bullet\sigma_\bullet)$ in the homoscedastic case $\varepsilon_\bullet = (\varepsilon)_{j\in\mathbb{N}}$, $\sigma_\bullet = (\sigma)_{j\in\mathbb{N}}$.

a_j	v_j	$k_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet)$ $\theta^\circ_\bullet \in \ell^2$	$\rho^2_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet)$	$k_{a_\bullet,\mathrm{v}_\bullet}(\theta^\circ_\bullet\sigma_\bullet)$ $\theta^\circ_\bullet = (j^{-t})_{j\in\mathbb{N}}$	$\rho^2_{a_\bullet,\mathrm{v}_\bullet}(\theta^\circ_\bullet\sigma_\bullet)$	
j^{-s}	j^{-p}	$\varepsilon^{-\frac{4}{4p+4s+1}}$	$\varepsilon^{\frac{8s}{4s+4p+1}}$	$\sigma^{-\frac{4}{4s+4(p-t)+1}}$	$\sigma^{\frac{8s}{4s+4(p-t)+1}}$	$t-p<\frac{1}{4}$
				$\sigma^{-\frac{1}{s}}$	$\lvert\log\sigma\rvert^{\frac{1}{2}}\,\sigma^2$	$t-p=\frac{1}{4}$
				$\sigma^{-\frac{1}{s}}$	σ^2	$t-p>\frac{1}{4}$
j^{-s}	e^{-j^p}	$\lvert\log\varepsilon\rvert^{\frac{1}{p}}$	$\lvert\log\varepsilon\rvert^{-\frac{2s}{p}}$	$\lvert\log\sigma\rvert^{\frac{1}{p}}$	$\lvert\log\sigma\rvert^{-\frac{2s}{p}}$	
e^{-j^s}	j^{-p}	$\lvert\log\varepsilon\rvert^{\frac{1}{s}}$	$\varepsilon^2\,\lvert\log\varepsilon\rvert^{\frac{4p+1}{2s}}$	$\lvert\log\sigma\rvert^{\frac{1}{s}}$	$\lvert\log\sigma\rvert^{\frac{4(p-t)+1}{2s}}\,\sigma^2$	$t-p<\frac{1}{4}$
				$\lvert\log\sigma\rvert^{\frac{1}{s}}$	$(\log\lvert\log\sigma\rvert)^{\frac{1}{2}}\,\sigma^2$	$t-p=\frac{1}{4}$
				$\lvert\log\sigma\rvert^{\frac{1}{s}}$	σ^2	$t-p>\frac{1}{4}$

Remark 1.2.6 (Simplified test statistics). *Let us note that by applying Markov's inequality it can be shown that the test* $\mathbb{1}_{\{\hat{\mathrm{q}}^2_{k_\star} > \tilde{\tau}_k(\alpha)\}}$ *with the simplified threshold* $\tilde{\tau}_k(\alpha) := \sqrt{\frac{2}{\alpha}}\mathrm{q}_{k_\star}\left(\frac{(\varepsilon^\circ_\bullet)^2}{\mathrm{v}^2_\bullet}\right)$ *and* $k_\star$ *as in* (1.2.9) *also attains the minimax radius of testing* $\rho_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet) \vee \rho_{a_\bullet,\mathrm{v}_\bullet}(\theta^\circ_\bullet\sigma_\bullet)$*. The approach of deriving radii of testing by applying Markov's inequality has for example been used in Kroll [2019a] and is used in Chapter 3 of this thesis. Since we are in particular concerned with adaptive Bonferroni aggregation, we need the sharper bound given in Proposition 1.2.1 for the threshold constant in terms of* α*. This directly translates to the cost to pay for adaptivity.* □

The test $\Delta_{k,\alpha}$ in (1.2.8) explicitly uses the knowledge of $\mathrm{v}_\bullet$, which determines the asymptotic behaviour of the sequence $\lambda_\bullet \in \Lambda^{\mathrm{c}}_{\mathrm{v}_\bullet}$. Inspired by Laurent et al. [2011], as an alternative we consider a direct test in Section 1.4. But, first, we provide a matching lower bound to the upper bound derived in Theorem 1.2.2 for the case $\varepsilon_\bullet \leq \sigma_\bullet$. This assumption is discussed in Remark 1.2.3.

1.3 Lower bound

Proposition 1.3.1 (Lower bound in terms of $\varepsilon_\bullet$). Let $k_\star := k_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet)$ and let $\eta \in (0,1]$ satisfy

$$\eta \leq \frac{\mathrm{q}_{k_\star}(\varepsilon^2_\bullet/\mathrm{v}^2_\bullet) \wedge a^2_{k_\star}}{\rho^2_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet)} = \frac{\mathrm{q}_{k_\star}(\varepsilon^2_\bullet/\mathrm{v}^2_\bullet) \wedge a^2_{k_\star}}{\mathrm{q}_{k_\star}(\varepsilon^2_\bullet/\mathrm{v}^2_\bullet) \vee a^2_{k_\star}}. \tag{1.3.1}$$

For $\alpha \in (0,1)$ define $\underline{A}^2_\alpha := \eta\left(\mathrm{R}^2 \wedge \sqrt{2\log(1+2\alpha^2)}\right)$. Then, for all $A \leq \underline{A}_\alpha$ we have

$$\mathcal{R}\left(\Theta^{\mathrm{R}}_{a_\bullet}, \Lambda^{\mathrm{c}}_{\mathrm{v}_\bullet}, \theta^\circ_\bullet, A\rho_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet)\right) \geq 1-\alpha,$$

i.e. $\rho^2_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet)$ is a lower bound for the minimax radius of testing.

Proof of Proposition 1.3.1. **Reduction step.** To prove lower bounds for the testing radius we reduce the risk of a test to a distance between probability measures on the null and the alternative. Let us write $\rho := \rho_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet)$ and let μ be a probability measure on $\left\{\ell^2_{\underline{A}_\alpha\rho} \cap \Theta^{\mathrm{R}}_{a_\bullet}\right\} \times \Lambda^{\mathrm{c}}_{\mathrm{v}_\bullet}$,

μ induces a so-called mixing measure

$$\mathbb{P}^{\varepsilon_\bullet,\sigma_\bullet}_{\mu} := \int_{\Theta^{\mathrm{R}}_{a_\bullet}\times\Lambda^{\mathrm{c}}_{\mathrm{v}_\bullet}} \mathbb{P}^{\varepsilon_\bullet,\sigma_\bullet}_{\theta_\bullet,\lambda_\bullet}\mathrm{d}\mu(\theta_\bullet,\lambda_\bullet).$$

For any test Δ the risk can then be lower bounded in terms of the χ^2-distance between the probability distribution under the null $\mathbb{P}_0 := \mathbb{P}^{\varepsilon_\bullet,\sigma_\bullet}_{\theta^\circ_\bullet,\mathrm{v}_\bullet}$ for $\mathrm{v}_\bullet \in \Lambda^{\mathrm{c}}_{\mathrm{v}_\bullet}$ and $\mathbb{P}_\mu := \mathbb{P}^{\varepsilon_\bullet,\sigma_\bullet}_{\mu}$ as follows

$$\begin{aligned}\mathcal{R}\left(\Delta \mid \Theta^{\mathrm{R}}_{a_\bullet},\Lambda^{\mathrm{c}}_{\mathrm{v}_\bullet},\theta^\circ_\bullet,\underline{A}_\alpha\rho\right) &\geq \mathbb{P}_0\left(\Delta=1\right)+\mathbb{P}_\mu\left(\Delta=0\right)\\ &= 1-\mathrm{TV}(\mathbb{P}_0,\mathbb{P}_\mu) \geq 1-\sqrt{\frac{\chi^2(\mathbb{P}_0,\mathbb{P}_\mu)}{2}},\end{aligned}\tag{1.3.2}$$

where TV denotes the total variation distance and χ^2 the χ^2-divergence.

Definition of the mixture. Let $k\in\mathbb{N}$ be fixed, for a given sequence of deviations from the null $\tilde{\theta}_\bullet \in \ell^2$ and $\tau\in\{\pm\}^k$ we define $\tilde{\theta}^\tau_\bullet\in\ell^2$ by

$$\widetilde{\theta}^\tau_j = \tau_j\widetilde{\theta}_j\mathbb{1}_{\{j\in[\![k]\!]\}}$$

We consider the uniform mixture measure over the vertices of a hypercube

$$\mathbb{P}_\mu := \frac{1}{2^k}\sum_{\tau\in\{\pm\}^k}\mathbb{P}^{\varepsilon_\bullet,\sigma_\bullet}_{\theta^\circ_\bullet+\tilde{\theta}^\tau_\bullet,\mathrm{v}_\bullet} = \frac{1}{2^k}\sum_{\tau\in\{\pm\}^k}\mathcal{N}(\mathrm{v}_\bullet(\theta^\circ_\bullet+\tilde{\theta}^\tau_\bullet),\varepsilon_\bullet)\otimes\mathcal{N}(\mathrm{v}_\bullet,\sigma_\bullet).$$

Naturally, since we only mix over $\Theta^{\mathrm{R}}_{a_\bullet}$ and not over $\Lambda^{\mathrm{c}}_{\mathrm{v}_\bullet}$, the χ^2-divergence between $\mathbb{P}_0$ and $\mathbb{P}_\mu$ reduces to the χ^2-divergence between the marginal distribution of $(Y_j)_{j\in\mathbb{N}}$ and the dependence on the marginal distribution of $(X_j)_{j\in\mathbb{N}}$ cancels. Lemma A.3.1 from the appendix then shows that

$$\begin{aligned}\chi^2(\mathbb{P}_0,\mathbb{P}_\mu) &= \chi^2\left(\frac{1}{2^k}\sum_{\tau\in\{\pm\}^k}\mathcal{N}(\mathrm{v}_\bullet(\theta^\circ_\bullet+\tilde{\theta}^\tau_\bullet),\varepsilon_\bullet),\mathcal{N}(\mathrm{v}_\bullet\theta^\circ_\bullet,\varepsilon_\bullet)\right)\\ &\leq \exp\left(\frac{1}{2}\sum_{j\in[\![k]\!]}\frac{\mathrm{v}_j^4\widetilde{\theta}_j^4}{\varepsilon_j^4}\right)-1 = \exp\left(\frac{1}{2}\mathrm{q}_k^2\left(\frac{\mathrm{v}_\bullet^2\widetilde{\theta}_\bullet^2}{\varepsilon_\bullet^2}\right)\right)-1.\end{aligned}$$

Combining the last bound with (1.3.2) we see that the assertion follows as soon as

(a) $\widetilde{\theta}\in\ell^2_{\underline{A}_\alpha\rho}$, (separation)

(b) $\widetilde{\theta}\in\Theta^{\mathrm{R}}_{a_\bullet}$, (smoothness)

(c) $\mathrm{q}_k^2\left(\frac{\mathrm{v}_\bullet^2\tilde{\theta}_\bullet^2}{\varepsilon_\bullet^2}\right)\leq 2\log(1+2\alpha^2)$. (similarity)

Definition of the deviations. It remains to define these quantities. Let $k := k_\star := k_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet)$ and consider $\tilde{\theta}_\bullet$ with

$$\widetilde{\theta}_j := \frac{\sqrt{\zeta\eta}\rho_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet)}{\mathrm{q}_{k_\star}(\varepsilon_\bullet^2/\mathrm{v}_\bullet^2)}\frac{\varepsilon_j^2}{\mathrm{v}_j^2}\mathbb{1}_{\{j\in[\![k_\star]\!]\}},\qquad \text{for } j\in\mathbb{N} \text{ and } \zeta := \mathrm{R}^2\wedge\sqrt{2\log(1+2\alpha^2)}.$$

Since $\left\|\tilde{\theta}_\bullet\right\|^2_{\ell^2} = \mathrm{q}^2_{k_\star}(\tilde{\theta}_\bullet) = \zeta\eta\rho^2_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet)$ with $\underline{A}_\alpha^2 = \zeta\eta$, (a) is satisfied. Moreover, the condition on η implies for all $m\leq k_\star$ that $\mathrm{b}^2_m(\tilde{\theta}_\bullet)\leq\mathrm{q}^2_{k_\star}(\tilde{\theta}_\bullet) = \underline{A}_\alpha^2\rho^2\leq\zeta a^2_{k_\star}\leq \mathrm{R}^2 a_m^2$ due to the monotonicity of $a_\bullet$. Trivially, we also have $\mathrm{b}^2_m(\tilde{\theta}_\bullet) = 0\leq a_m^2$ for each $m>k_\star$. Therefore, $\tilde{\theta}_\bullet$ satisfies (b). Again exploiting the condition on η we obtain

$$\mathrm{q}^2_{k_\star}\left(\frac{\mathrm{v}_\bullet^2\tilde{\theta}_\bullet^2}{\varepsilon_\bullet^2}\right) = \zeta^2\eta^2\frac{(\rho_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet))^4}{\mathrm{q}^2_{k_\star}(\varepsilon_\bullet^2/\mathrm{v}_\bullet^2)}\leq\zeta^2\leq 2\log(1+2\alpha^2),$$

and, thus, also (c) holds, which completes the proof. □

Note that the lower bound in Proposition 1.3.1 involves the value η satisfying (1.3.1), which depends on the joint behaviour of the sequences $\mathrm{v}_\bullet$ and $a_\bullet$ and essentially guarantees an optimal balance of the bias and the variance term in the dimension $k_\star$. Moreover, looking at Remark 1.2.3 we see that Proposition 1.3.1 provides a matching lower bound to the upper bound derived in Theorem 1.2.2, whenever the radius is governed by the $\varepsilon_\bullet$-term, which is for instance the case if $\varepsilon_\bullet \geq \sigma_\bullet$.

1.4 Upper bound via a direct testing procedure

In this section we derive an upper bound for the radius of testing based on the estimation of $\|\lambda_\bullet(\theta_\bullet - \theta^\circ_\bullet)\|^2_{\ell^2}$ instead of $\left\|\frac{\lambda_\bullet}{\mathrm{v}_\bullet}(\theta_\bullet - \theta^\circ_\bullet)\right\|^2_{\ell^2}$ as in the section before. In fact, for $k \in \mathbb{N}$ we consider

$$\tilde{\mathrm{q}}^2_k := \sum_{j\in[\![k]\!]} \left((Y_j - \theta^\circ_j X_j)^2 - (\varepsilon^\circ_j)^2\right) = \sum_{j\in[\![k]\!]} \left(\tilde{Y}^2_j - (\varepsilon^\circ_j)^2\right), \tag{1.4.1}$$

which is an unbiased estimator of the truncated version $\mathrm{q}^2_k(\lambda_\bullet(\theta_\bullet - \theta^\circ_\bullet))$, since

$$\mathbb{E}^{\varepsilon_\bullet,\sigma_\bullet}_{\theta_\bullet,\lambda_\bullet}\left(\tilde{\mathrm{q}}^2_k\right) = \sum_{|j|\in[\![k]\!]} \left((\varepsilon^\circ_j)^2 + \lambda^2_j\left(\theta_j - \theta^\circ_j\right)^2 - (\varepsilon^\circ_j)^2\right) = \mathrm{q}^2_k(\lambda_\bullet(\theta_\bullet - \theta^\circ_\bullet)).$$

To formulate a result similar to Proposition 1.2.1 we introduce for a sequence $x_\bullet \in \mathbb{R}^{\mathbb{N}}$ the minimum

$$\left(\rho^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(x_\bullet)\right)^2 := \min_{k\in\mathbb{N}}\left\{\mathrm{v}_k^{-2}\mathrm{q}_k\left(x^2_\bullet\right) \vee a^2_k\right\} = \min_{k\in\mathbb{N}}\left\{\mathrm{v}_k^{-2}\sqrt{\sum_{j\in[\![k]\!]} x^4_j} \vee a^2_k\right\} \tag{1.4.2}$$

and minimizer

$$k^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(x_\bullet) := \arg\min_{k\in\mathbb{N}}\left\{\mathrm{v}_k^{-2}\mathrm{q}_k\left(x^2_\bullet\right) \vee a^2_k\right\} = \arg\min_{k\in\mathbb{N}}\left\{\mathrm{v}_k^{-2}\sqrt{\sum_{j\in[\![k]\!]} x^4_j} \vee a^2_k\right\}. \tag{1.4.3}$$

Replacing the sequence $x_\bullet$ by the **original** and **effective** noise levels $\varepsilon_\bullet$ and $\theta^\circ_\bullet\sigma_\bullet$ we establish $(\rho^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet) \vee \rho^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\theta^\circ_\bullet\sigma_\bullet))^2$ as the optimal achievable radius for the direct test. Similar to Proposition 1.2.1 (for the indirect test) the next result allows to evaluate the performance of the direct test based on the test statistic (1.4.1) under both, the null hypothesis and the alternative.

Proposition 1.4.1 (Bounds for the quantiles of $\tilde{\mathrm{q}}^2_k$). For $u \in (0,1)$ set $L_u := \sqrt{|\log u|}$. Let $\alpha, \beta \in (0,1)$.

(i) **(α-level)** For each $k \in \mathbb{N}$ we have

$$\sup_{\lambda\in\Lambda^{\mathrm{c}}_{\mathrm{v}_\bullet}} \mathbb{P}^{\varepsilon_\bullet,\sigma_\bullet}_{\theta^\circ_\bullet,\lambda_\bullet}\left(\tilde{\mathrm{q}}^2_k > 2L_\alpha \mathrm{q}^2_k\left((\varepsilon^\circ_\bullet)^2\right) + 2L^2_\alpha \mathrm{m}_k\left((\varepsilon^\circ_\bullet)^2\right)\right) \leq \alpha. \tag{1.4.4}$$

(ii) **($(1-\beta)$-powerful)** Define the dimension $k^{\mathrm{d}}_\star := k^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet) \wedge k^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\theta^\circ_\bullet\sigma_\bullet) \in \mathbb{N}$ as in (1.2.3) and $C_{\alpha,\beta} := 5\left(L_\alpha + L^2_\alpha + L_\beta + 5L^2_\beta\right)$. Then for each $\theta_\bullet - \theta^\circ_\bullet \in \Theta^{\mathrm{R}}_{a_\bullet} \cap \ell^2_\rho$ with

$$\rho^2 \geq \left(\mathrm{R}^2 + \mathrm{c}C_{\alpha,\beta}\right)(\rho^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet) \vee \rho^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\theta^\circ_\bullet\sigma_\bullet))^2$$

we have

$$\sup_{\lambda\in\Lambda^{c}_{\mathrm{v}_\bullet}} \mathbb{P}^{\varepsilon_\bullet;\sigma_\bullet}_{\theta_\bullet,\lambda_\bullet}\left(\hat{\mathrm{q}}^2_{k^{\mathrm{d}}_\star} \le 2L_\alpha \mathrm{q}^2_{k^{\mathrm{d}}_\star}\left((\varepsilon^\circ_\bullet)^2\right) + 2L_\alpha^2 \mathrm{m}_{k^{\mathrm{d}}_\star}\left((\varepsilon^\circ_\bullet)^2\right)\right) \le \beta. \tag{1.4.5}$$

Proof of Proposition 1.4.1. We note that $(Y_\bullet, X_\bullet) \sim \mathbb{P}^{\varepsilon_\bullet;\sigma_\bullet}_{\theta_\bullet,\lambda_\bullet}$ implies

$$\mathrm{Q}_k := \widetilde{\mathrm{q}}^2_k + \mathrm{q}^2_k(\varepsilon^\circ_\bullet) \sim \mathbb{Q}^{e_\bullet}_{\mu_\bullet,k}$$

with $e_\bullet = \varepsilon^\circ_\bullet$ and $\mu_\bullet = \lambda_\bullet(\theta_\bullet - \theta^\circ_\bullet)$, where we again use the notation of Lemma A.1.1 in the appendix.

(i) The proof of (1.4.4) follows analogously to the proof of (1.2.4) in Proposition 1.2.1 by applying Lemma A.1.1.

(ii) Similar calculations as in the proof of (1.2.5) show that for each $\theta_\bullet - \theta^\circ_\bullet \in \Theta^{\mathrm{R}}_{a_\bullet} \cap \ell^2_\rho$ with $\rho^2 \ge (\mathrm{R}^2 + \mathrm{c}C_{\alpha,\beta})\left(\rho^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet) \vee \rho^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\theta^\circ_\bullet\sigma_\bullet)\right)^2$, we obtain for $k = k^{\mathrm{d}}_\star$

$$\begin{aligned}
\|\theta_\bullet - \theta^\circ_\bullet\|^2_{\ell^2} &\ge \rho^2 \ge (\mathrm{R}^2 + \mathrm{c}C_{\alpha,\beta})\left(\rho^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet) \vee \rho^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\theta^\circ_\bullet\sigma_\bullet)\right)^2 \\
&\ge \mathrm{R}^2 a_k^2 + \mathrm{c}C_{\alpha,\beta}\mathrm{v}_k^{-2}\left\{\mathrm{q}_k\left(\varepsilon^2_\bullet\right) \vee \mathrm{q}_k\left((\theta^\circ_\bullet)^2\sigma^2_\bullet\right)\right\} \\
&\ge \mathrm{R}^2 a_k^2 + \mathrm{c}\frac{C_{\alpha,\beta}}{2}\mathrm{v}_k^{-2}\mathrm{q}_k\left(e^2_\bullet\right) \\
&\ge \mathrm{R}^2 a_k^2 + \mathrm{c}\frac{5}{2}\mathrm{v}_k^{-2}\left\{L_\alpha \mathrm{q}_k\left(e^2_\bullet\right) + L^2_\alpha \mathrm{m}_k\left(e^2_\bullet\right) + \mathrm{q}_k\left(e^2_\bullet\right)(L_\beta + 5L^2_\beta)\right\},
\end{aligned} \tag{1.4.6}$$

using $\left(\rho^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet) \vee \rho^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\theta^\circ_\bullet\sigma_\bullet)\right)^2 = \mathrm{v}_k^{-2}\mathrm{q}_k(\varepsilon^2_\bullet) \vee \mathrm{q}_k((\theta^\circ_\bullet)^2\sigma^2_\bullet) \vee a_k^2$ due to Lemma A.2.1 and $2\{\mathrm{q}_k(\varepsilon^2_\bullet) \vee \mathrm{q}_k((\theta^\circ_\bullet)^2\sigma^2_\bullet)\} \ge \mathrm{q}_k((\varepsilon^\circ_\bullet)^2) = \mathrm{q}_k(e^2_\bullet) \ge \mathrm{m}_k(e^2_\bullet)$. Moreover, for each $k \in \mathbb{N}$ and $\lambda \in \Lambda^{c}_{\mathrm{v}_\bullet}$ we have $\mathrm{c}\mathrm{v}_k^{-2}\mathrm{q}^2_k(\mu_\bullet) \ge \mathrm{q}^2_k(\theta_\bullet - \theta^\circ_\bullet) = \|\theta_\bullet - \theta^\circ_\bullet\|^2_{\ell^2} - \mathrm{b}^2_k(\theta_\bullet - \theta^\circ_\bullet)$, which in turn for each $\theta_\bullet - \theta^\circ_\bullet \in \Theta^{\mathrm{R}}_{a_\bullet}$ implies

$$\mathrm{c}\mathrm{v}_k^{-2}\mathrm{q}^2_k(\mu_\bullet) \ge \|\theta_\bullet - \theta^\circ_\bullet\|^2_{\ell^2} - \mathrm{R}^2 a_k^2.$$

This bound applied with $k = k^{\mathrm{d}}_\star$ together with (1.4.6) yields

$$\mathrm{c}\mathrm{v}_k^{-2}\mathrm{q}^2_k(\mu_\bullet) \ge \mathrm{c}\frac{5}{2}\mathrm{v}_k^{-2}\left\{L_\alpha \mathrm{q}_k\left(e^2_\bullet\right) + L^2_\alpha \mathrm{m}_k\left(e^2_\bullet\right) + \mathrm{q}_k\left(e^2_\bullet\right)(L_\beta + 5L^2_\beta)\right\}.$$

Rearranging the last inequality and proceeding as in the proof of (1.2.5) implies (1.4.5), which shows the assertion.

□

Definition of the test. For $\alpha \in (0,1)$ and $k \in \mathbb{N}$ we define the threshold

$$\tau^{\mathrm{d}}_k(\alpha) := 2L_\alpha \mathrm{q}^2_k((\varepsilon^\circ_\bullet)^2) + 2L^2_\alpha \mathrm{m}_k((\varepsilon^\circ_\bullet)^2) \tag{1.4.7}$$

and the corresponding test

$$\Delta^{\mathrm{d}}_{k,\alpha} := \mathbb{1}_{\left\{\widetilde{\mathrm{q}}^2_k > \tau^{\mathrm{d}}_k(\alpha)\right\}}. \tag{1.4.8}$$

Proposition 1.4.1 (i) shows that the test $\Delta^{\mathrm{d}}_{k,\alpha/2}$ is a level $\alpha/2$-test for any $k \in \mathbb{N}$. Moreover, $\Delta^{\mathrm{d}}_{k^{\mathrm{d}}_\star,\alpha/2}$ with

$$k^{\mathrm{d}}_\star := k^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet) \wedge k^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\theta^\circ_\bullet\sigma_\bullet) \tag{1.4.9}$$

is a $(1-\alpha/2)$-powerful test over $\overline{A}_\alpha \left\{\rho^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet) \vee \rho^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\theta^\circ_\bullet\sigma_\bullet)\right\}$-separated alternatives due to Proposition 1.4.1 (ii) with $\beta = \alpha/2$ and $\overline{A}^2_\alpha := \mathrm{R}^2 + \mathrm{c}C_{\alpha/2,\alpha/2} = \mathrm{R}^2 + \mathrm{c}\left(10L_{\alpha/2} + 30L^2_{\alpha/2}\right)$. Hence,

$$\mathcal{R}\left(\Delta^{\mathrm{d}}_{k^{\mathrm{d}}_\star,\alpha/2} \mid \Theta^{\mathrm{R}}_{a_\bullet}, \Lambda^{\mathrm{c}}_{\mathrm{v}_\bullet}, \theta^\circ_\bullet, A\left\{\rho^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet) \vee \rho^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\theta^\circ_\bullet\sigma_\bullet)\right\}\right) \leq \alpha/2 + \alpha/2 = \alpha$$

for all $A \geq \overline{A}_\alpha$. In other words, $\left(\rho^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet) \vee \rho^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\theta^\circ_\bullet\sigma_\bullet)\right)^2$ is an upper bound for the radius of testing of $\Delta^{\mathrm{d}}_{k^{\mathrm{d}}_\star,\alpha/2}$. Moreover, it is also a lower bound for its radius of testing, which we prove in the next proposition.

Proposition 1.4.2 (Radius of testing of $\Delta^{\mathrm{d}}_{k^{\mathrm{d}}_\star,\alpha/2}$). Let $\alpha \in (0,1)$ and

$$\rho := \rho^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet) \vee \rho^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\theta^\circ_\bullet\sigma_\bullet).$$

(i) (**upper bound**) With $\overline{A}^2_\alpha := \mathrm{R}^2 + \mathrm{c}\left(10L_{\alpha/2} + 30L^2_{\alpha/2}\right)$ we obtain for all $A \geq \overline{A}_\alpha$

$$\mathcal{R}\left(\Delta^{\mathrm{d}}_{k^{\mathrm{d}}_\star,\alpha/2} \mid \Theta^{\mathrm{R}}_{a_\bullet}, \Lambda^{\mathrm{c}}_{\mathrm{v}_\bullet}, \theta^\circ_\bullet, A\rho\right) \leq \alpha.$$

(ii) (**lower bound**) Let

$$0 < \eta \leq \frac{a^2_{k^D_\star}}{\rho^2}$$

and define $\underline{A}^2_\alpha := \mathrm{R}^2\eta$. Then it follows for all $A \leq \underline{A}_\alpha$

$$\mathcal{R}\left(\Delta^{\mathrm{d}}_{k^{\mathrm{d}}_\star,\alpha/2} \mid \Theta^{\mathrm{R}}_{a_\bullet}, \Lambda^{\mathrm{c}}_{\mathrm{v}_\bullet}, \theta^\circ_\bullet, A\rho\right) \geq 1 - \alpha.$$

Summarizing, ρ^2 is a radius of testing for the test $\Delta^{\mathrm{d}}_{k^{\mathrm{d}}_\star,\alpha/2}$.

Proof of Proposition 1.4.2. Firstly, part (i) is an immediate consequence of Proposition 1.4.1 and we omit the details. Secondly, consider part (ii). We note that for each $\lambda_\bullet \in \Lambda^{\mathrm{c}}_{\mathrm{v}_\bullet}$ and $\theta_\bullet - \theta^\circ_\bullet \in \Theta^{\mathrm{R}}_{a_\bullet}$ with $\mathrm{q}^2_{k^{\mathrm{d}}_\star}(\theta_\bullet - \theta^\circ_\bullet) = 0$ we have

$$\mathrm{Q}_{k^{\mathrm{d}}_\star} := \tilde{\mathrm{q}}^2_{k^{\mathrm{d}}_\star} + \mathrm{q}^2_{k^{\mathrm{d}}_\star}(\varepsilon^\circ_\bullet) \sim \mathbb{Q}^{\varepsilon^\circ_\bullet}_{0_\bullet,k^{\mathrm{d}}_\star}$$

and, thus,

$$\mathbb{P}^{\varepsilon_\bullet,\sigma_\bullet}_{\theta_\bullet,\lambda_\bullet}(\Delta^{\mathrm{d}}_{k^{\mathrm{d}}_\star,\alpha/2} = 1) \leq \frac{\alpha}{2}$$

due to (A.1.1) in Lemma A.1.1 (using the notation introduced therein). For any $\theta_\bullet - \theta^\circ_\bullet \in \Theta^{\mathrm{R}}_{a_\bullet}$ with

$$\mathrm{q}^2_{k^{\mathrm{d}}_\star}(\theta_\bullet - \theta^\circ_\bullet) = 0 \qquad \text{and} \qquad \mathrm{b}^2_{k^{\mathrm{d}}_\star}(\theta_\bullet - \theta^\circ_\bullet) = \mathrm{R}^2 a^2_{k^{\mathrm{d}}_\star},$$

which is for instance satisfied for

$$\theta_\bullet - \theta^\circ_\bullet = \left(\mathrm{R}a_{k^{\mathrm{d}}_\star}\mathbb{1}_{\{j=k^{\mathrm{d}}_\star+1\}}\right)_{j\in\mathbb{N}},$$

it immediately follows

$$\underline{A}_\alpha^2\rho^2 = \mathrm{R}^2\eta\left(\rho^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet)\vee\rho^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\theta^\circ_\bullet\sigma_\bullet)\right)^2 \leq \mathrm{R}^2 a^2_{k^{\mathrm{d}}_\star}$$
$$= \mathrm{b}^2_{k^{\mathrm{d}}_\star}(\theta_\bullet-\theta^\circ_\bullet) \leq \|\theta_\bullet-\theta^\circ_\bullet\|^2_{\ell^2},$$

which shows that $\theta_\bullet - \theta^\circ_\bullet$ is contained in the alternative. Hence, for such a $\theta_\bullet$ and all $A \geq \underline{A}_\alpha$ we obtain

$$\mathcal{R}\left(\Delta^{\mathrm{d}}_{k^{\mathrm{d}}_\star,\alpha/2} \mid \Theta^{\mathrm{R}}_{a_\bullet},\Lambda^{\mathrm{c}}_{\mathrm{v}_\bullet},\theta^\circ_\bullet,A\rho\right) \geq \mathbb{P}^{\varepsilon_\bullet,\sigma_\bullet}_{\theta_\bullet,\lambda_\bullet}(\Delta^{\mathrm{d}}_{k^{\mathrm{d}}_\star,\alpha/2}=0) \geq 1-\frac{\alpha}{2} \geq 1-\alpha,$$

which shows (ii) and completes the proof. □

Remark 1.4.3 (Optimality of the direct test). *Considering the* ***signal detection task****, i.e.* $\theta^\circ_\bullet = 0_\bullet$ *we have* $\rho^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\theta^\circ_\bullet\sigma_\bullet) = 0$ *for all* $\sigma_\bullet \in \mathbb{R}^{\mathbb{N}}_+$ *and thus the radius of testing does not depend on the noise level* $\sigma_\bullet$*. Considering the* ***goodness-of-fit task****, i.e.* $\theta^\circ_\bullet \neq 0_\bullet$*, we emphasise that for all* $\varepsilon_\bullet \geq \sigma_\bullet$*, we have*

$$\mathrm{q}_k\left((\theta^\circ_\bullet)^2\sigma^2_\bullet\right) \leq \mathrm{q}_k\left((\theta^\circ_\bullet)^2\varepsilon^2_\bullet\right) \leq \|\theta^\circ_\bullet\|^2_{\ell^\infty}\,\mathrm{q}_k\left(\varepsilon^2_\bullet\right)$$

and, therefore, $\rho^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\theta^\circ_\bullet\sigma_\bullet) \leq \|\theta^\circ_\bullet\|_{\ell^\infty}\,\rho^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet)$*. In other words,* $\rho^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\theta^\circ_\bullet\sigma_\bullet)$ *is negligible compared to* $\|\theta^\circ_\bullet\|_{\ell^\infty}\,\rho^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet)$*. Hence, the direct test shows a similar behaviour as the indirect test (as discussed in Remark 1.2.3).*

Let us now briefly discuss under which conditions the direct test attains the radius $\rho^2_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet)\vee\rho^2_{a_\bullet,\mathrm{v}_\bullet}(\theta^\circ_\bullet\sigma_\bullet)$ *of the indirect test. For any* $\varepsilon_\bullet \in \mathbb{R}^{\mathbb{N}}_+$ *the elementary inequality*

$$\mathrm{v}_k^{-2}\mathrm{q}_k^2(\varepsilon^2_\bullet) \geq \mathrm{q}_k^2\left(\frac{\varepsilon^2_\bullet}{\mathrm{v}^2_\bullet}\right), \qquad k \in \mathbb{N} \tag{1.4.10}$$

shows that

$$\rho^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet) \geq \rho_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet).$$

If there exists a constant $\kappa_1 \in \mathbb{R}_+$ *such that*

$$\mathrm{v}_k^{-2}\mathrm{q}_k^2(\varepsilon^2_\bullet) \leq \kappa_1\mathrm{q}_k^2\left(\frac{\varepsilon^2_\bullet}{\mathrm{v}^2_\bullet}\right), \qquad k \in \mathbb{N}, \tag{1.4.11}$$

then $\rho^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet)$ *and* $\rho_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet)$ *are of the same order. In particular, then the test* $\Delta^{\mathrm{d}}_{k^{\mathrm{d}}_\star,\alpha/2}$ *attains the minimax radius in the signal detection case. The condition (1.4.11) is e.g. satisfied in a* ***mildly ill-posed*** *model. Note that, however, the additional condition is sufficient but not necessary as we will see in the illustration below. Considering the radius in terms of* $\sigma_\bullet$ *we obtain* $\rho^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\theta^\circ_\bullet\sigma_\bullet) \geq \rho_{a_\bullet,\mathrm{v}_\bullet}(\theta^\circ_\bullet\sigma_\bullet)$ *by exploiting again the elementary inequality (1.4.10) (with* $\varepsilon_\bullet$ *replaced by* $\theta^\circ_\bullet\sigma_\bullet$*). Therefore, if there exists in addition a constant* $\kappa_2 \in \mathbb{R}_+$ *such that*

$$\mathrm{v}_k^{-2}\mathrm{q}_k^2((\theta^\circ_\bullet)^2\sigma^2_\bullet) \leq \kappa_2\mathrm{q}_k^2\left(\frac{(\theta^\circ_\bullet)^2\sigma^2_\bullet}{\mathrm{v}^2_\bullet}\right), \qquad k \in \mathbb{N}, \tag{1.4.12}$$

then $\rho^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\theta^\circ_\bullet\sigma_\bullet)$ *and* $\rho_{a_\bullet,\mathrm{v}_\bullet}(\theta^\circ_\bullet\sigma_\bullet)$ *are of the same order. Summarizing, if both (1.4.11) and (1.4.12) are satisfied, then the test* $\Delta^{\mathrm{d}}_{k^{\mathrm{d}}_\star,\alpha/2}$ *attains the same radius as the indirect test, where the conditions are again sufficient but not necessary.* □

Illustration 1.4.4. In the homoscedastic case we illustrate the order of the rate and the corresponding optimal dimension parameter of the direct test $\Delta^{\mathrm{d}}_{k^{\mathrm{d}}_{\star},\alpha/2}$ defined in (1.4.8) by considering the typical smoothness and ill-posedness assumptions as in Illustration 1.2.5. The table displays the order of the rate $\left(\rho^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet)\vee\rho^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\theta^\circ_\bullet\sigma_\bullet)\right)^2$ for the signal detection task $\theta^\circ_\bullet = 0_\bullet$ (with $\rho^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\theta^\circ_\bullet\sigma_\bullet) = 0$ as discussed in Remark 1.4.3) and the goodness-of-fit task with $\theta^\circ_\bullet = (j^{-t})_{j\in\mathbb{N}}$. In comparison with Illustration 1.2.5 we point out that in all three cases the orders of $\left(\rho^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet)\right)^2$ and $\rho^2_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet)$ coincide. Note that there exists a $\kappa_1\in\mathbb{R}_+$ such that (1.4.11) is fulfilled only in the case of a mildly ill-posed model. In a severely ill-posed model, however, there exists no such constant. Nonetheless, in both cases the direct test performs optimally with respect to the noise level ε. Comparing the orders of $\left(\rho^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\theta^\circ_\bullet\sigma_\bullet)\right)^2$ and $\rho^2_{a_\bullet,\mathrm{v}_\bullet}(\theta^\circ_\bullet\sigma_\bullet)$ we note that in both a mildly and a severely ill-posed model there does not exist a $\kappa_2\in\mathbb{R}_+$ such that (1.4.12) is satisfied. Even so, for severely ill-posed models the rate of the direct test $\left(\rho^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet)\vee\rho^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\theta^\circ_\bullet\sigma_\bullet)\right)^2$ and the rate of the indirect test $\rho^2_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet)\vee\rho^2_{a_\bullet,\mathrm{v}_\bullet}(\theta^\circ_\bullet\sigma_\bullet)$ are of the same order, and thus the direct test is also optimal. On the other hand, for mildly ill-posed models the rate $\left(\rho^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\theta^\circ_\bullet\sigma_\bullet)\right)^2$ is always nonparametric and might be much slower than the rate $\rho^2_{a_\bullet,\mathrm{v}_\bullet}(\theta^\circ_\bullet\sigma_\bullet)$, which can be parametric (cp. Illustration 1.2.5).

Order of the optimal dimension $k^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet)\wedge k^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\theta^\circ_\bullet\sigma_\bullet)$
and the radius $\left(\rho^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet)\vee\rho^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\theta^\circ_\bullet\sigma_\bullet)\right)^2$
in the homoscedastic case $\varepsilon_\bullet = (\varepsilon)_{j\in\mathbb{N}}$ and $\sigma_\bullet = (\sigma)_{j\in\mathbb{N}}$

a_j (smooth.)	v_j (ill-posed.)	$k^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet)$ $\theta^\circ_\bullet\in\ell^2$	$\left(\rho^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet)\right)^2$	$k^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\theta^\circ_\bullet\sigma_\bullet)$ $\theta^\circ_\bullet=(j^{-t})_{j\in\mathbb{N}}$	$\left(\rho^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\theta^\circ_\bullet\sigma_\bullet)\right)^2$
j^{-s}	j^{-p}	$\varepsilon^{-\frac{4}{4p+4s+1}}$	$\varepsilon^{\frac{8s}{4s+4p+1}}$	$\sigma^{-\frac{1}{s+p}}$	$\sigma^{\frac{2s}{s+p}}$
j^{-s}	e^{-j^p}	$\lvert\log\varepsilon\rvert^{\frac{1}{p}}$	$\lvert\log\varepsilon\rvert^{-\frac{2s}{p}}$	$\lvert\log\sigma\rvert^{\frac{1}{p}}$	$\lvert\log\sigma\rvert^{-\frac{2s}{p}}$
e^{-j^s}	j^{-p}	$\lvert\log\varepsilon\rvert^{\frac{1}{s}}$	$\varepsilon^2\lvert\log\varepsilon\rvert^{\frac{4p+1}{2s}}$	$\lvert\log\sigma\rvert^{\frac{1}{s}}$	$\lvert\log\sigma\rvert^{\frac{2p}{s}}\sigma^2$

Laurent et al. [2011] show that for known operators, under specific smoothness and ill-posedness assumptions (covered also in Illustration 1.2.5 and Illustration 1.4.4) every minimax optimal test for the direct task is also minimax optimal for the indirect task. Even under these specific assumptions this in no longer the case for unknown operators if $\rho^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet)$ is negligible compared to $\rho^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\theta^\circ_\bullet\sigma_\bullet)$, since we observe that the direct test $\Delta^{\mathrm{d}}_{k^{\mathrm{d}}_{\star},\alpha/2}$ defined in (1.4.8) is not always optimal for the indirect task (compare Illustration 1.4.4 and Illustration 1.2.5).

1.5 Adaptation

1.5.1 Description of the adaptation procedure

For both the indirect and the direct test the optimal choice of the dimension parameter $k_\star$ (in (1.2.3)) respectively $k^{\mathrm{d}}_\star$ (in (1.4.3)) require prior knowledge of the sequences $a_\bullet$ and $\mathrm{v}_\bullet$, which are typically unknown in practise. In this section we study an aggregation of the tests over several dimension parameters, which leads to a testing procedure that performs nearly optimal over a wide range of regularity classes. We first present the testing radii of these aggregation-tests, where compared to the minimax radii of testing we observe a deterioration by a log-factor with respect to the noise levels. Moreover, we show that this deterioration is an unavoidable cost to pay for adaptation.

Aggregation procedure. Let us briefly describe a widely used aggregation strategy. For a sequence of levels $(\alpha_k)_{k\in\mathbb{N}} \in (0,1)^{\mathbb{N}}$ let $(S_{k,\alpha_k})_{k\in\mathbb{N}}$ be a sequence of test statistics such that

$$\nabla_{k,\alpha_k} := \mathbb{1}_{\{S_{k,\alpha_k}>0\}}$$

is a level-α_k-test for each $k \in \mathbb{N}$. Note that both the indirect and the direct testing procedures satisfy this condition by construction as shown in (1.2.4) and (1.4.4) in Proposition 1.2.1 and Proposition 1.4.1, respectively. Given a finite collection $\mathcal{K} \subseteq \mathbb{N}$ of dimension parameters and $\alpha := \sum_{k\in\mathcal{K}} \alpha_k$ we consider the max-test statistic

$$S_{\mathcal{K},\alpha} := \max_{k\in\mathcal{K}} S_{k,\alpha_k}$$

and the max-test

$$\nabla_{\mathcal{K},\alpha} := \mathbb{1}_{\{S_{\mathcal{K},\alpha}>0\}},$$

that is, the max-test rejects the null hypothesis as soon as one of the tests does. Due to the elementary inequality

$$\mathbb{P}^{\varepsilon_\bullet;\sigma_\bullet}_{\theta^\circ_\bullet,\lambda_\bullet}(\nabla_{\mathcal{K},\alpha}=1) = \mathbb{P}^{\varepsilon_\bullet;\sigma_\bullet}_{\theta^\circ_\bullet,\lambda_\bullet}(S_{\mathcal{K},\alpha}>0) \leq \sum_{k\in\mathcal{K}} \mathbb{P}^{\varepsilon_\bullet;\sigma_\bullet}_{\theta^\circ_\bullet,\lambda_\bullet}(S_{k,\alpha_k}>0) \leq \sum_{k\in\mathcal{K}} \alpha_k = \alpha, \tag{1.5.1}$$

the max-test $\nabla_{\mathcal{K},\alpha}$ is a level-α-test. The type II error probability of the max-test can be controlled by any test contained in the collection, since for all $\theta_\bullet \in \ell^2$ and $\lambda_\bullet \in \ell^\infty$ we have

$$\mathbb{P}^{\varepsilon_\bullet;\sigma_\bullet}_{\theta_\bullet,\lambda_\bullet}(\nabla_{\mathcal{K},\alpha}=0) = \mathbb{P}^{\varepsilon_\bullet;\sigma_\bullet}_{\theta_\bullet,\lambda_\bullet}(S_{\mathcal{K},\alpha}\leq 0) \leq \min_{k\in\mathcal{K}} \mathbb{P}^{\varepsilon_\bullet;\sigma_\bullet}_{\theta_\bullet,\lambda_\bullet}(S_{k,\alpha_k}\leq 0) = \min_{k\in\mathcal{K}} \mathbb{P}^{\varepsilon_\bullet;\sigma_\bullet}_{\theta_\bullet,\lambda_\bullet}(\nabla_{k,\alpha_k}=0).$$

These two error bounds have oppositional effects on the choice of the collection $\mathcal{K}$. Roughly speaking, $\mathcal{K}$ should not be too large due to the aggregation of type I error probabilities. On the other hand it should still be large enough to minimise the type II error probabilities. Typically the choice of $\mathcal{K}$ depends on the original and the effective noise levels $\varepsilon_\bullet$ and $\theta^\circ_\bullet\sigma_\bullet$.

Bonferroni correction. Throughout this section, given a level $\alpha \in (0,1)$ and a finite collection $\mathcal{K} \subseteq \mathbb{N}$ we consider Bonferroni levels $\alpha_k = \frac{\alpha}{|\mathcal{K}|}$, $k \in \mathcal{K}$, i.e. the same level $\frac{\alpha}{|\mathcal{K}|}$ for each test statistic S_{k,α_k} in the collection. For alternative constructions of error levels we refer to Remark 4.3.1, where they are discussed in detail in a similar setting.

Lack of adaptability. The goal of the aggregation is to find testing strategies for which the risk can be controlled over large families of alternatives. Let $\mathcal{A} \subseteq \ell^2$ and $\mathcal{V} \subseteq \ell^\infty$ be classes of positive, monotonically non-increasing sequences that are bounded by 1. To measure the cost to pay for adaptation we introduce factors $\delta_{\varepsilon_\bullet}$ and $\delta_{\sigma_\bullet}$, which are typically called **adaptive factors** (cf. Spokoiny [1996]) for a family of tests $\{\nabla_\alpha\}_{\alpha\in(0,1)}$ and a family of alternatives $\left\{\Theta^{\mathrm{R}}_{a_\bullet} : a_\bullet \in \mathcal{A}\right\} \times \{\Lambda^{\mathrm{c}}_{\mathrm{v}_\bullet} : \mathrm{v}_\bullet \in \mathcal{V}\}$, if for every $\alpha \in (0,1)$ there exists a constant $\overline{A}_\alpha \in \mathbb{R}_+$ such that for all $A \geq \overline{A}_\alpha$ we have

$$\sup_{(a_\bullet,\mathrm{v}_\bullet)\in\mathcal{A}\times\mathcal{V}} \mathcal{R}\left(\nabla_\alpha \mid \Theta^{\mathrm{R}}_{a_\bullet}, \Lambda^{\mathrm{c}}_{\mathrm{v}_\bullet}, \theta^\circ_\bullet, A\left\{\rho_{a_\bullet,\mathrm{v}_\bullet}(\delta_{\varepsilon_\bullet}\varepsilon_\bullet) \vee \rho_{a_\bullet,\mathrm{v}_\bullet}(\delta_{\sigma_\bullet}\theta^\circ_\bullet\sigma_\bullet)\right\}\right) \leq \alpha.$$

Here, $\rho_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet) \vee \rho_{a_\bullet,\mathrm{v}_\bullet}(\theta^\circ_\bullet\sigma_\bullet)$ is a non-adaptive radius of testing over $\Theta^{\mathrm{R}}_{a_\bullet} \times \Lambda^{\mathrm{c}}_{\mathrm{v}_\bullet}$. Compared with the non-adaptive upper bound, the test ∇_α now needs to perform for any combination

$(a_\bullet, \mathrm{v}_\bullet) \in \mathcal{A} \times \mathcal{V}$. We, however, allow larger radii where the noise levels $\varepsilon_\bullet$ and $\theta^\circ_\bullet \sigma_\bullet$ are magnified by the factors $\delta_{\varepsilon_\bullet}, \delta_{\sigma_\bullet} > 1$. The factors $\delta_{\varepsilon_\bullet}$ and $\delta_{\sigma_\bullet}$ are called **minimal adaptive factors** if in addition for every $\alpha \in (0,1)$ there exists a constant $\underline{A}_\alpha$ such that for all $A \leq \underline{A}_\alpha$

$$\inf_{\nabla} \sup_{(a_\bullet, \mathrm{v}_\bullet) \in \mathcal{A} \times \mathcal{V}} \mathcal{R}\left(\nabla \mid \Theta^{\mathrm{R}}_{a_\bullet}, \Lambda^{\mathrm{c}}_{\mathrm{v}_\bullet}, \theta^\circ_\bullet, A\left\{\rho_{a_\bullet, \mathrm{v}_\bullet}(\delta_{\varepsilon_\bullet}\varepsilon_\bullet) \vee \rho_{a_\bullet, \mathrm{v}_\bullet}(\delta_{\sigma_\bullet}\theta^\circ_\bullet\sigma_\bullet)\right\}\right) \geq 1 - \alpha.$$

If the minimal adaptive factors tend to infinity as the noise levels decrease to zero, then this phenomenon is typically called **lack of adaptability**.

1.5.2 Adaptation to smoothness – indirect test

In this section we first carry out an aggregation of the indirect tests. Recall that the indirect test statistic $\hat{\mathrm{q}}_k^2$ in (1.2.1) and the threshold $\tau_k(\alpha)$ in (1.2.7) explicitly use the knowledge of the sequence $\mathrm{v}_\bullet$. Therefore, we consider adaptation to $\left\{\Theta^{\mathrm{R}}_{a_\bullet} : a_\bullet \in \mathcal{A}\right\} \times \Lambda^{\mathrm{c}}_{\mathrm{v}_\bullet}$ for a given $\mathrm{v}_\bullet$ only. We present the adaptive factors for the indirect max-test and show that they coincide with the minimal adaptive factors asymptotically.

Indirect max-**test.** Given $\alpha \in (0,1)$ and a finite collection $\mathcal{K} \subseteq \mathbb{N}$ we define the max-test statistic with Bonferroni levels

$$T_{\mathcal{K},\alpha} := \max_{k \in \mathcal{K}} \left\{\hat{\mathrm{q}}_k^2 - \tau_k\left(\frac{\alpha}{|\mathcal{K}|}\right)\right\}$$

and the corresponding test

$$\Delta_{\mathcal{K},\alpha} := \mathbb{1}_{\{T_{\mathcal{K},\alpha} > 0\}},$$

which is a level-α-test due to (1.2.4) in Proposition 1.2.1 and (1.5.1). Its radius of testing faces a deterioration compared with the minimax radius due to the Bonferroni aggregation, which we formalise next. Analogously to (1.2.2), for each sequence $x_\bullet \in \mathbb{R}^{\mathbb{N}}$ we define the minimum over the collection $\mathcal{K}$

$$\rho^2_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(x_\bullet) := \min_{k \in \mathcal{K}} \left\{\mathrm{q}_k\left(\frac{x_\bullet^2}{\mathrm{v}_\bullet^2}\right) \vee a_k^2\right\} = \min_{k \in \mathcal{K}} \left\{\sqrt{\sum_{j \in [\![k]\!]} \frac{x_j^4}{\mathrm{v}_j^4}} \vee a_k^2\right\} \tag{1.5.2}$$

and the corresponding minimizer

$$k_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(x_\bullet) := \arg\min_{k \in \mathcal{K}} \left\{\mathrm{q}_k\left(\frac{x_\bullet^2}{\mathrm{v}_\bullet^2}\right) \vee a_k^2\right\} = \arg\min_{k \in \mathcal{K}} \left\{\sqrt{\sum_{j \in [\![k]\!]} \frac{x_j^4}{\mathrm{v}_j^4}} \vee a_k^2\right\}. \tag{1.5.3}$$

Additionally, we define the minimum

$$r^2_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(x_\bullet) := \min_{k \in \mathcal{K}} \left\{\mathrm{m}_k\left(\frac{x_\bullet^2}{\mathrm{v}_\bullet^2}\right) \vee a_k^2\right\} = \min_{k \in \mathcal{K}} \left\{\max_{j \in [\![k]\!]} \frac{x_j^2}{\mathrm{v}_j^2} \vee a_k^2\right\}. \tag{1.5.4}$$

We first provide an upper bound for the radius of testing of the max-test in terms of the reparametrised noise level $(\varepsilon^\circ_\bullet)^2 = \varepsilon_\bullet^2 + (\theta^\circ_\bullet)^2\sigma_\bullet^2$ and the adaptive factor

$$\delta_{\mathcal{K}} := (1 \vee \log|\mathcal{K}|)^{1/4}. \tag{1.5.5}$$

The upper bound consists of the maximum of two terms

$$\rho^2_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\mathcal{K}}\varepsilon^\circ_\bullet) \qquad \text{and} \qquad r^2_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta^2_{\mathcal{K}}\varepsilon^\circ_\bullet).$$

We think of $r^2_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta^2_{\mathcal{K}}\varepsilon^\circ_\bullet)$ as a remainder term, which is typically negligible compared with $\rho^2_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\mathcal{K}}\varepsilon^\circ_\bullet)$ (compare Remark 1.5.2 below).

Proposition 1.5.1 (Adaptive upper bound – indirect max-test). For $\alpha \in (0,1)$ define $\overline{A}_\alpha^2 := \mathrm{R}^2 + \mathrm{c}\left(5L_{\alpha/2} + 15L_{\alpha/2}^2 + 5\right)$. Then for all $A \geq \overline{A}_\alpha$ we obtain

$$\sup_{a_\bullet \in \mathcal{A}} \mathcal{R}\left(\Delta_{\mathcal{K},\alpha/2} \mid \Theta^{\mathrm{R}}_{a_\bullet}, \Lambda^{\mathrm{c}}_{\mathrm{v}_\bullet}, \theta^\circ_\bullet, A\rho_{a_\bullet}\right) \leq \alpha.$$

with $\rho_{a_\bullet} = \rho_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\mathcal{K}}\varepsilon^\circ_\bullet) \vee r_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta^2_{\mathcal{K}}\varepsilon^\circ_\bullet)$ and $\delta_{\mathcal{K}} := (1 \vee \log|\mathcal{K}|)^{1/4}$

Proof of Proposition 1.5.1. The proof follows along the lines of the proof of Proposition 1.2.1 and exploits (A.1.1) and (A.1.3) in Lemma A.1.1. We use the notation introduced there. For $(Y_\bullet, X_\bullet) \sim \mathbb{P}^{\varepsilon_\bullet;\sigma_\bullet}_{\theta_\bullet,\lambda_\bullet}$ we have

$$\mathrm{Q}_k := \mathrm{q}_k^2 + \mathrm{q}_k^2(e_\bullet) \sim \mathbb{Q}^{e_\bullet}_{\mu_\bullet,k}$$

for each $k \in \mathbb{N}$ and $e^2_\bullet := \frac{(\varepsilon^\circ_\bullet)^2}{\mathrm{v}^2_\bullet}$ and $\mu_\bullet := \frac{\lambda_\bullet}{\mathrm{v}_\bullet}(\theta_\bullet - \theta^\circ_\bullet)$. (A.1.1) implies that **under the null hypothesis** with $L := \sqrt{\log(2|\mathcal{K}|/\alpha)}$ the quantile satisfies

$$\mathrm{q}^{e_\bullet}_{0_\bullet,k}\left(\frac{\alpha}{2|\mathcal{K}|}\right) \leq \mathrm{q}_k^2(e_\bullet) + 2L\mathrm{q}_k(e^2_\bullet) + 2L^2\mathrm{m}_k(e^2_\bullet)$$

and, therefore,

$$\begin{aligned}\mathbb{P}^{\varepsilon_\bullet;\sigma_\bullet}_{\theta^\circ_\bullet,\lambda_\bullet}\left(\Delta_{\mathcal{K},\alpha/2} = 1\right) &= \mathbb{P}^{\varepsilon_\bullet;\sigma_\bullet}_{\theta^\circ_\bullet,\lambda_\bullet}\left(T_{\mathcal{K},\alpha} > 0\right) \leq \sum_{k\in\mathcal{K}} \mathbb{P}^{\varepsilon_\bullet;\sigma_\bullet}_{\theta^\circ_\bullet,\lambda_\bullet}\left(\hat{\mathrm{q}}_k^2 > \tau_k(\alpha/|\mathcal{K}|)\right)\\ &= \sum_{k\in\mathcal{K}} \mathbb{P}^{\varepsilon_\bullet;\sigma_\bullet}_{\theta^\circ_\bullet,\lambda_\bullet}\left(\mathrm{Q}_k > \mathrm{q}^{e_\bullet}_{0_\bullet,k}\right) \leq \sum_{k\in\mathcal{K}} \frac{\alpha}{2|\mathcal{K}|} = \frac{\alpha}{2}. \end{aligned} \tag{1.5.6}$$

Under the alternative for $\theta_\bullet - \theta^\circ_\bullet \in \Theta^{\mathrm{R}}_{a_\bullet} \cap \ell^2_\rho$ with $\rho \geq \overline{A}_\alpha \left\{\rho_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\mathcal{K}}\varepsilon^\circ_\bullet) \vee r_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta^2_{\mathcal{K}}\varepsilon^\circ_\bullet)\right\}$ we have

$$\begin{aligned}\|\theta_\bullet - \theta^\circ_\bullet\|^2_{\ell^2} &\geq \overline{A}^2_\alpha\left\{\rho^2_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\mathcal{K}}\varepsilon^\circ_\bullet) \vee r^2_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta^2_{\mathcal{K}}\varepsilon^\circ_\bullet)\right\}\\ &\geq \mathrm{R}^2 a^2_{k_\star} + \mathrm{c}\left(5L_{\alpha/2} + 15L^2_{\alpha/2} + 5\right)\left\{\mathrm{q}_{k_\star}\left(\delta^2_{\mathcal{K}}e^2_\bullet\right) \vee \mathrm{m}_{k_\star}(\delta^4_{\mathcal{K}}e^2_\bullet)\right\}\\ &\geq \mathrm{R}^2 a^2_{k_\star} + \mathrm{c}5L_{\alpha/2}\mathrm{q}_{k_\star}\left(\delta^2_{\mathcal{K}}e^2_\bullet\right) + \mathrm{c}15L^2_{\alpha/2}\mathrm{m}_{k_\star}(\delta^4_{\mathcal{K}}e^2_\bullet) + \mathrm{c}5\mathrm{q}_{k_\star}\left(\delta^2_{\mathcal{K}}e^2_\bullet\right)\\ &\geq \mathrm{R}^2 a^2_{k_\star} + \mathrm{c}\frac{5}{2}\left(L\mathrm{q}_{k_\star}\left(e^2_\bullet\right) + L^2\mathrm{m}_{k_\star}\left(e^2_\bullet\right) + \mathrm{q}^2_{k_\star}\left(e^2_\bullet\right)\left(L_{\alpha/2} + 5L^2_{\alpha/2}\right)\right)\end{aligned} \tag{1.5.7}$$

where we successively use $(*)$, $(**)$ and $(***)$ shown below. Indeed, we have

$$\rho^2_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\mathcal{K}}\varepsilon^\circ_\bullet) \vee r^2_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta^2_{\mathcal{K}}\varepsilon^\circ_\bullet) = \mathrm{q}_{k_\star}\left(\delta^2_{\mathcal{K}}e^2_\bullet\right) \vee \mathrm{m}_{k_\star}(\delta^4_{\mathcal{K}}e^2_\bullet) \vee a^2_{k_\star} \tag{$*$}$$

with $k_\star := \arg\min_{k\in\mathcal{K}}\{\mathrm{m}_k(\delta^4_{\mathcal{K}}e^2_\bullet)\} \wedge \arg\min_{k\in\mathcal{K}}\{\mathrm{q}_k(\delta^2_{\mathcal{K}}e^2_\bullet) \vee a^2_k\}$ due to Lemma A.2.1;

$$\mathrm{q}_{k_\star}(\delta^2_{\mathcal{K}}e^2_\bullet) \geq \mathrm{q}_{k_\star}(e^2_\bullet) \tag{$**$}$$

and

$$\mathrm{q}_{k_\star}(\delta^2_{\mathcal{K}}e^2_\bullet)(L_{\alpha/2}+1) \geq L\mathrm{q}_{k_\star}(e^2_\bullet), \qquad \mathrm{m}_{k_\star}(\delta^4_{\mathcal{K}}e^2_\bullet)(L^2_{\alpha/2}+1) \geq L^2\mathrm{m}_{k_\star}(e^2_\bullet). \tag{$***$}$$

For all $k \in \mathbb{N}$, $\lambda_\bullet \in \Lambda^{\mathrm{c}}_{\mathrm{v}_\bullet}$ and $\theta_\bullet - \theta^\circ_\bullet \in \Theta^{\mathrm{R}}_{a_\bullet} \cap \ell^2_\rho$ it follows

$$\mathrm{c}\mathrm{q}^2_k(\mu_\bullet) \geq \|\theta_\bullet - \theta^\circ_\bullet\|^2_{\ell^2} - \mathrm{R}^2 a^2_k,$$

which together with (1.5.7) implies

$$\frac{4}{5}\mathrm{q}_{k_\star}^2(\mu_\bullet) \geq 2L\mathrm{q}_{k_\star}(e_\bullet^2) + 2L^2\mathrm{m}_{k_\star}(e_\bullet^2) + 2(L_{\alpha/2} + 5L_{\alpha/2}^2)\mathrm{q}_{k_\star}(e_\bullet^2).$$

Rearranging the last inequality and using (A.1.3) in Lemma A.1.1 shows that for all $\lambda_\bullet \in \Lambda_{\mathrm{v}_\bullet}^{\mathrm{c}}$ under the alternative we have

$$\begin{aligned}\mathbb{P}_{\theta_\bullet,\lambda_\bullet}^{\varepsilon_\bullet,\sigma_\bullet}\left(\Delta_{\mathcal{K},\alpha/2} = 0\right) &= \mathbb{P}_{\theta_\bullet,\lambda_\bullet}^{\varepsilon_\bullet,\sigma_\bullet}\left(T_{\mathcal{K},\alpha/2} \leq 0\right)\\ &\leq \min_{k\in\mathcal{K}} \mathbb{P}_{\theta_\bullet,\lambda_\bullet}^{\varepsilon_\bullet,\sigma_\bullet}\left(\mathbb{Q}_k \leq 2L\mathrm{q}_k(e_\bullet^2) + 2L^2\mathrm{m}_k(e_\bullet^2) + \mathrm{q}_k^2(e_\bullet)\right)\\ &\leq \mathbb{P}_{\theta_\bullet,\lambda_\bullet}^{\varepsilon_\bullet,\sigma_\bullet}\left(\mathbb{Q}_{k_\star} \leq \mathrm{q}_{\mu_\bullet,k_\star}^{e_\bullet}(1-\alpha/2)\right) = \frac{\alpha}{2}. \end{aligned} \tag{1.5.8}$$

Combining the bound for the type I error probability (1.5.6) and the bound for the type II error probability (1.5.8) completes the proof. □

Remark 1.5.2 (Adaptive factor $\delta_{\mathcal{K}}^2$ versus $\delta_{\mathcal{K}}$). *The second term $r_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\mathcal{K}}^2\varepsilon_\bullet^\circ)$ in the upper bound of Proposition 1.5.1 for the adaptive radius of testing can always be bounded by $\rho_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\mathcal{K}}^2\varepsilon_\bullet^\circ)$ due to the elementary inequality $\mathrm{m}_k(\delta_{\mathcal{K}}^4(\varepsilon_\bullet^\circ)^2/\mathrm{v}_\bullet^2) \leq \mathrm{q}_k(\delta_{\mathcal{K}}^4(\varepsilon_\bullet^\circ)^2/\mathrm{v}_\bullet^2)$ for all $k \in \mathbb{N}$. Note that $\rho_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\mathcal{K}}^2\varepsilon_\bullet^\circ)$ only differs from the first term $\rho_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\mathcal{K}}\varepsilon_\bullet^\circ)$ in the upper bound of Proposition 1.5.1 by an additional factor $\delta_{\mathcal{K}}$. Hence, we can always show that $\delta_{\mathcal{K}}^2$* ***is an adaptive factor****. However, often this bound is too rough and the term $r_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\mathcal{K}}^2\varepsilon_\bullet^\circ)$ is negligible compared to $\rho_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\mathcal{K}}\varepsilon_\bullet^\circ)$, which then results in* ***an adaptive factor*** *$\delta_{\mathcal{K}}$. Let us give sufficient conditions for the negligibility. Consider $k_\star := k_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\mathcal{K}}\varepsilon_\bullet^\circ)$ as in (1.5.3). We give a condition in terms of the relationship between $k_\star$ and $\delta_{\mathcal{K}}$, which is then easy to check. Assume there exists a $C > 0$ such that*

$$\sqrt{k_\star}\mathrm{m}_{k_\star}\left(\frac{(\varepsilon_\bullet^\circ)^2}{\mathrm{v}_\bullet^2}\right) \leq C\mathrm{q}_{k_\star}\left(\frac{(\varepsilon_\bullet^\circ)^2}{\mathrm{v}_\bullet^2}\right), \tag{1.5.9}$$

i.e. we "gain" at least a factor $\sqrt{k_\star}$ by considering the maximum instead of the quadratic functional (this is for instance the case in a mildly ill-posed model with homogeneous variance) and assume additionally (for all $a_\bullet \in \mathcal{A}$)

$$\delta_{\mathcal{K}}^2 \leq C\sqrt{k_\star}. \tag{1.5.10}$$

Then, naturally

$$\mathrm{m}_{k_\star}\left(\frac{\delta_{\mathcal{K}}^4(\varepsilon_\bullet^\circ)^2}{\mathrm{v}_\bullet^2}\right) \leq C\sqrt{k_\star}\mathrm{m}_{k_\star}\left(\frac{\delta_{\mathcal{K}}^2(\varepsilon_\bullet^\circ)^2}{\mathrm{v}_\bullet^2}\right) \leq C^2\mathrm{q}_{k_\star}\left(\frac{\delta_{\mathcal{K}}^2(\varepsilon_\bullet^\circ)^2}{\mathrm{v}_\bullet^2}\right),$$

and, hence, $r_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\mathcal{K}}^2\varepsilon_\bullet^\circ) \leq C\rho_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\mathcal{K}}\varepsilon_\bullet^\circ)$, which implies that we obtain an adaptive factor $\delta_{\mathcal{K}}$. □

We now reformulate the upper bound in Proposition 1.5.1 in terms of the noise levels $\varepsilon_\bullet$ and $\theta_\bullet^\circ\sigma_\bullet$. Recall that the optimal dimension is given by a minimum

$$k_\star := k_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet) \wedge k_{a_\bullet,\mathrm{v}_\bullet}(\theta_\bullet^\circ\sigma_\bullet)$$

(compare Proposition 1.2.1 (ii)). Therefore, we eventually choose collections $\mathcal{K}_{\varepsilon_\bullet}$ and $\mathcal{K}_{\sigma_\bullet}$ depending on $\varepsilon_\bullet$ respectively $\sigma_\bullet$ only and set

$$\mathcal{K} := \mathcal{K}_{\varepsilon_\bullet} \cap \mathcal{K}_{\sigma_\bullet}$$

with

$$\delta_{\varepsilon_\bullet} := \delta_{\mathcal{K}_{\varepsilon_\bullet}} \qquad \text{and} \qquad \delta_{\sigma_\bullet} := \delta_{\mathcal{K}_{\sigma_\bullet}}.$$

Trivially, $|\mathcal{K}| \leq |\mathcal{K}_{\varepsilon_\bullet}| \wedge |\mathcal{K}_{\sigma_\bullet}|$ and hence $\delta_{\mathcal{K}} \leq \delta_{\varepsilon_\bullet} \wedge \delta_{\sigma_\bullet}$. The next result is a direct consequence of Proposition 1.5.1 due to $2\{\rho_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\varepsilon_\bullet}\varepsilon_\bullet) \vee \rho_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\sigma_\bullet}\theta_\bullet^\circ\sigma_\bullet)\} \geq \rho_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\mathcal{K}}\varepsilon_\bullet^\circ)$ and its proof is omitted. The assumption (1.5.11) simply states that the remainder term is indeed negligible.

Theorem 1.5.3 (Adaptive upper bound - indirect max-test). Let $\mathcal{K} := \mathcal{K}_{\varepsilon_\bullet} \cap \mathcal{K}_{\sigma_\bullet}$, $\delta_{\varepsilon_\bullet} := \delta_{\mathcal{K}_{\varepsilon_\bullet}}$ and $\delta_{\sigma_\bullet} := \delta_{\mathcal{K}_{\sigma_\bullet}}$. Assume there exists a $C \geq 1$ such that

$$r_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\mathcal{K}}^2 \varepsilon_\bullet^\circ) \leq C \rho_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\mathcal{K}} \varepsilon_\bullet^\circ) \tag{1.5.11}$$

for all $a_\bullet \in \mathcal{A}$. Then, for each $\alpha \in (0,1)$ with $\overline{A}_\alpha^2 := 2\left(\mathrm{R}^2 + \mathrm{c}\left(5L_{\alpha/2} + 15L_{\alpha/2}^2 + 5\right)\right)$ it follows for all $A \geq \overline{A}_\alpha$

$$\sup_{a_\bullet \in \mathcal{A}} \mathcal{R}\left(\Delta_{\mathcal{K},\alpha/2} \mid \Theta_{a_\bullet}^{\mathrm{R}}, \Lambda_{\mathrm{v}_\bullet}^{\mathrm{c}}, \theta_\bullet^\circ, AC\rho_{a_\bullet}\right) \leq \alpha$$

with $\rho_{a_\bullet} = \rho_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\varepsilon_\bullet}\varepsilon_\bullet) \vee \rho_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\sigma_\bullet}\theta_\bullet^\circ\sigma_\bullet)$.

In Remark 1.5.4 and Illustration 1.5.5 we select a suitable collection $\mathcal{K}$ such that the minimisation over $\mathcal{K}$ approximates the minimisation over $\mathbb{N}$ well, i.e. such that the upper bound in Theorem 1.5.3 satisfies

$$\rho_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\varepsilon_\bullet}\varepsilon_\bullet) \vee \rho_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\sigma_\bullet}\theta_\bullet^\circ\sigma_\bullet) \leq \tilde{C}\left\{\rho_{a_\bullet,\mathrm{v}_\bullet}(\delta_{\varepsilon_\bullet}\varepsilon_\bullet) \vee \rho_{a_\bullet,\mathrm{v}_\bullet}(\delta_{\sigma_\bullet}\theta_\bullet^\circ\sigma_\bullet)\right\}$$

for some $\tilde{C} > 1$.

Remark 1.5.4 (Choice of $\mathcal{K}$ in the homoscedastic setting). *Let us discuss the choice of the collection $\mathcal{K}$ of dimension parameters in the homoscedastic case $\varepsilon_\bullet = (\varepsilon)_{j\in\mathbb{N}}$, $\sigma_\bullet = (\sigma)_{j\in\mathbb{N}}$. Considering the* ***signal detection task*** *(where only the ε-terms appear, i.e. we set $\mathcal{K}_{\sigma_\bullet} = \mathbb{N}$ and $\mathcal{K} = \mathcal{K}_{\varepsilon_\bullet}$) it is easily seen that for all $\mathrm{v}_\bullet \in \mathcal{V}$ and $a_\bullet \in \mathcal{A}$ the minimax optimal dimension $k_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon_\bullet)$ is never larger than ε^{-4}. Therefore, the natural choice $\mathcal{K} = \mathcal{K}_\varepsilon := [\![\varepsilon^{-4}]\!]$ yields a factor $\delta_{\mathcal{K}}$ of order $|\log\varepsilon|^{1/4}$. However, in many cases it is sufficient to aggregate over a geometric grid $\mathcal{K}_g := \{2^j : j \in [\![4\,|\log_2\varepsilon|\,]\!]\} \cup \{1\}$. Obviously, $\delta_{\mathcal{K}_g}$ is then of order $(\log|\log\varepsilon|)^{1/4}$.*
For a ***goodness-of-fit task*** *the upper bound for the minimax optimal dimension parameter can further be improved by exploiting the knowledge of $\theta_\bullet^\circ$. More precisely, since $\mathrm{q}_k^2\left(\frac{(\varepsilon_\bullet^\circ)^2}{\mathrm{v}_\bullet}\right) \geq \mathrm{q}_k^2\left((\varepsilon_\bullet^\circ)^2\right) \geq \varepsilon^4 k + \sigma^4\mathrm{q}_k^2((\theta_\bullet^\circ)^2)$ and $\sup_{k\in\mathbb{N}} a_k \leq 1$, any $k \in \mathbb{N}$ such that $\sigma^4\mathrm{q}_k^2((\theta_\bullet^\circ)^2) \geq 1$ is an upper bound for the dimension parameter. For instance, for the goodness-of-fit task with $\theta_\bullet^\circ = (j^{-t})_{j\in\mathbb{N}}$ as considered in Illustration 1.5.5 below, the upper bound is of order σ^{-4}, which results in the natural choice $\mathcal{K} = [\![\varepsilon^{-4}]\!] \cap [\![\sigma^{-4}]\!] =: \mathcal{K}_\varepsilon \cap \mathcal{K}_\sigma$ and an adaptive factor $|\log\varepsilon|^{1/4} \wedge |\log\sigma|^{1/4}$. However, since a geometric grid $\mathcal{K}_g := \{2^j : j \in [\![4\,|\log_2\varepsilon|\,]\!] \cap [\![4\,|\log_2\sigma|\,]\!]\} \cup \{1\}$ is again sufficient, $\delta_{\mathcal{K}}$ is of order $(\log|\log\varepsilon|)^{1/4} \wedge (\log|\log\sigma|)^{1/4} = \delta_\varepsilon \wedge \delta_\sigma$.* □

Summarizing, to obtain the desired upper bound $\rho_{a_\bullet,\mathrm{v}_\bullet}^2(\delta_{\varepsilon_\bullet}\varepsilon_\bullet) \vee \rho_{a_\bullet,\mathrm{v}_\bullet}^2(\delta_{\sigma_\bullet}\theta_\bullet^\circ\sigma_\bullet)$ from Theorem 1.5.3, there are two things to do. Firstly, construct the collection $\mathcal{K}$ such that minimization over $\mathcal{K}$ approximates minimization over $\mathbb{N}$ well. Secondly, show the negligibility of the remainder term, i.e. verify (1.5.11). This is done in the illustration below.

Illustration 1.5.5 (Homoscedastic case). Consider the smoothness and ill-posedness assumptions of Illustration 1.2.5. Define the geometric grid

$$\mathcal{K} := \mathcal{K}_g := \left\{2^j : j \in [\![4\,|\log_2\varepsilon|\,]\!] \cap [\![4\,|\log_2\sigma|\,]\!]\right\} \cup \{1\}$$

with an adaptive factor $\delta_{\mathcal{K}} \leq \delta_{\varepsilon_\bullet} \wedge \delta_{\sigma_\bullet}$, where $\delta_\varepsilon \sim (\log|\log\varepsilon|)^{1/4}$ and $\delta_\sigma \sim (\log|\log\sigma|)^{1/4}$. As discussed in Remark 1.5.4 **in all three cases** minimization over $\mathcal{K}$ approximates minimization over $\mathbb{N}$ well, i.e. there exists a $C > 0$ such that, uniformly for all $a_\bullet \in \mathcal{A}$,

$$C\rho_{a_\bullet,\mathrm{v}_\bullet}(\delta_{\mathcal{K}}\varepsilon^\circ_\bullet) \geq \rho_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\mathcal{K}}\varepsilon^\circ_\bullet).$$

Moreover, for **mildly ill-posed** models $r_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta^2_{\mathcal{K}}\varepsilon^\circ_\bullet)$ is negligible compared with $\rho_{a_\bullet,\mathrm{v}_\bullet}(\delta_{\mathcal{K}}\varepsilon^\circ_\bullet)$, i.e. uniformly for all $s \in [s_\star, s^\star]$ there exists a $\tilde{C} \geq 1$ such that

$$r_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta^2_{\mathcal{K}}\varepsilon^\circ_\bullet) \leq \tilde{C}\rho_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\mathcal{K}}\varepsilon^\circ_\bullet),$$

since the conditions (1.5.9) and (1.5.10) are fulfilled. Furthermore, the constant $\tilde{C}$ can be chosen uniformly for all sufficiently small noise levels.
In a **severely ill-posed** model $r_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta^2_{\mathcal{K}}\varepsilon^\circ_\bullet)$, $\rho_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\mathcal{K}}\varepsilon^\circ_\bullet)$, $\rho_{a_\bullet,\mathrm{v}_\bullet}(\delta_{\mathcal{K}}\varepsilon^\circ_\bullet)$ and $\rho_{a_\bullet,\mathrm{v}_\bullet}(\varepsilon^\circ_\bullet)$ are all of the same order and the adaptive factors have no effect on the rate. We present the resulting rates of testing $\rho^2_{a_\bullet,\mathrm{v}_\bullet}(\delta_{\varepsilon_\bullet}\varepsilon_\bullet) \vee \rho^2_{a_\bullet,\mathrm{v}_\bullet}(\delta_{\sigma_\bullet}\theta^\circ_\bullet\sigma_\bullet)$ for both the signal detection $(\theta^\circ_\bullet = 0_\bullet)$ and the goodness-of-fit task $(\theta^\circ_\bullet = (j^{-t})_{j\in\mathbb{N}})$ in the table below. Note that we only consider the case $4t - 4p < 1$ to avoid unnecessary case distinctions and increase the readability of the table.

Order of the adaptive radius $\rho^2_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\varepsilon_\bullet}\varepsilon_\bullet) \vee \rho^2_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\sigma_\bullet}\theta^\circ_\bullet\sigma_\bullet)$
for the geometric grid $\mathcal{K} := \mathcal{K}_g := \{2^j : j \in [\![4\,|\log_2\varepsilon|\,]\!] \cap [\![4\,|\log_2\sigma|\,]\!]\} \cup \{1\}$
in the homoscedastic case $\varepsilon_\bullet = (\varepsilon)_{j\in\mathbb{N}}$ and $\sigma_\bullet = (\sigma)_{j\in\mathbb{N}}$.

a_j (smooth.)	v_j (ill-posed.)	$\rho^2_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\varepsilon_\bullet}\varepsilon_\bullet)$ $\theta^\circ_\bullet \in \ell^2$	$\rho^2_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\sigma_\bullet}\theta^\circ_\bullet\sigma_\bullet)$ $\theta^\circ_\bullet = (j^{-t})_{j\in\mathbb{N}}$, $4t - 4p < 1$
j^{-s}	j^{-p}	$\Big((\log\lvert\log\varepsilon\rvert)^{\frac{1}{4}}\varepsilon\Big)^{\frac{8s}{4s+4p+1}}$	$((\log\lvert\log\sigma\rvert)^{\frac{1}{4}}\sigma)^{\frac{8s}{4s+4(p-t)+1}}$
j^{-s}	e^{-j^p}	$\lvert\log\varepsilon\rvert^{-\frac{2s}{p}}$	$\lvert\log\sigma\rvert^{-\frac{2s}{p}}$
e^{-j^s}	j^{-p}	$\varepsilon^2(\log\lvert\log\varepsilon\rvert)^{\frac{1}{2}}\,\lvert\log\varepsilon\rvert^{\frac{4p+1}{2s}}$	$\sigma^2(\log\lvert\log\sigma\rvert)^{\frac{1}{2}}\,\lvert\log\sigma\rvert^{\frac{4(p-t)+1}{2s}}$

In case of super smoothness $a_\bullet = (e^{-j^s})_{j\in\mathbb{N}}$ and mild ill-posedness (see Illustration 1.2.5) the minimax optimal dimension parameter is of order $|\log\varepsilon|^{\frac{1}{2s}}$ in the signal detection case and of order $|\log\varepsilon|^{\frac{1}{2s}} \wedge |\log\sigma|^{\frac{1}{2s}}$ in the goodness-of-fit task, which suggest (for adaptation to $s \geq s^\star$) a smaller geometric grid

$$\mathcal{K}_{s_\star} := \left\{2^j : j \in [\![\frac{1}{2s_\star}\log_2|\log\varepsilon|\,]\!] \cap [\![\frac{1}{2s_\star}\log_2|\log\sigma|\,]\!]\right\} \cup \{1\}\,,$$

yielding an adaptive factor $\delta_{\mathcal{K}_{s_\star}} \lesssim \delta_\varepsilon \wedge \delta_\sigma$ with $\delta_\varepsilon \sim (\log\log|\log\varepsilon|)^{\frac{1}{4}}$ and $\delta_\sigma \sim (\log\log|\log\sigma|)^{\frac{1}{4}}$. Indeed, in this situation there exists a $C \geq 1$ such that

$$r_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta^2_{\mathcal{K}}\varepsilon^\circ_\bullet) \vee \rho_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\mathcal{K}}\varepsilon^\circ_\bullet) \leq C\rho_{a_\bullet,\mathrm{v}_\bullet}(\delta_{\mathcal{K}}\varepsilon^\circ_\bullet)$$

uniformly for all $s \geq s_\star$ and for sufficiently small noise levels. We present the resulting rates of testing in the table below.

Order of the adaptive radius $\rho^2_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\varepsilon_\bullet}\varepsilon_\bullet) \vee \rho^2_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\sigma_\bullet}\theta^\circ_\bullet\sigma_\bullet)$ for the geometric grid $\mathcal{K} := \mathcal{K}_{s_\star} := \left\{2^j : j \in [\![\frac{1}{2s_\star}\log_2|\log\varepsilon|]\!] \cap [\![\frac{1}{2s_\star}\log_2|\log\sigma|]\!]\right\}$ in the homoscedastic case $\varepsilon_\bullet = (\varepsilon)_{j\in\mathbb{N}}$ and $\sigma_\bullet = (\sigma)_{j\in\mathbb{N}}$.			
a_j (smooth.)	v_j (ill-posed.)	$\rho^2_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\varepsilon_\bullet}\varepsilon_\bullet)$ $\theta^\circ_\bullet \in \ell^2$	$\rho^2_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\sigma_\bullet}\theta^\circ_\bullet\sigma_\bullet)$ $\theta^\circ_\bullet = (j^{-t})_{j\in\mathbb{N}}$, $4t-4p<1$
e^{-j^s}	j^{-p}	$\varepsilon^2(\log\log|\log\varepsilon|)^{\frac{1}{2}}\,|\log\varepsilon|^{\frac{4p+1}{2s}}$	$\sigma^2(\log\log|\log\sigma|)^{\frac{1}{2}}\,|\log\sigma|^{\frac{4(p-t)+1}{2s}}$

1.5.3 Adaptation to both smoothness and ill-posedness – direct test

As an alternative to the indirect test we have introduced the direct test in Section 1.4. In contrast to the indirect test it only depends on the sequence $\mathrm{v}_\bullet$ through the choice of the optimal dimension parameter. Hence, for a direct-max-test by aggregating over various dimension parameters we consider adaptation to both smoothness $\left\{\Theta^{\mathrm{R}}_{a_\bullet} : a_\bullet \in \mathcal{A}\right\}$ and ill-posedness $\{\Lambda^{\mathrm{c}}_{\mathrm{v}_\bullet} : \mathrm{v}_\bullet \in \mathcal{V}\}$.

Direct max-test. Given $\alpha \in (0,1)$ and a finite collection $\mathcal{K} \subseteq \mathbb{N}$ we define the max-test statistic with Bonferroni levels

$$T^{\mathrm{d}}_{\mathcal{K},\alpha} := \max_{k\in\mathcal{K}}\left\{\widetilde{\mathrm{q}}^2_k - \tau^{\mathrm{d}}_k\left(\frac{\alpha}{|\mathcal{K}|}\right)\right\}$$

and the corresponding test

$$\Delta^{\mathrm{d}}_{\mathcal{K},\alpha} := \mathbb{1}_{\left\{T^{\mathrm{d}}_{\mathcal{K},\alpha}>0\right\}},$$

which is a level-α-test due to (1.4.4) in Proposition 1.4.1. Its testing radius faces a deterioration compared to the optimal direct testing radius derived in Proposition 1.4.2 due to the Bonferroni aggregation. Analogously to (1.4.2), for a sequence $x_\bullet \in \mathbb{R}^{\mathbb{N}}$ we define a minimum over the collection $\mathcal{K}$

$$(\rho^{\mathrm{d}}_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(x_\bullet))^2 := \min_{k\in\mathcal{K}}\left\{\mathrm{v}_k^{-2}\mathrm{q}_k(x^2_\bullet) \vee a^2_k\right\} = \min_{k\in\mathcal{K}}\left\{\mathrm{v}_k^{-2}\sqrt{\sum_{j\in[\![k]\!]} x_j^4} \vee a^2_k\right\} \tag{1.5.12}$$

and the corresponding minimiser

$$k^{\mathrm{d}}_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(x_\bullet) := \arg\min_{k\in\mathcal{K}}\left\{\mathrm{v}_k^{-2}\mathrm{q}_k(x^2_\bullet) \vee a^2_k\right\} = \arg\min_{k\in\mathcal{K}}\left\{\mathrm{v}_k^{-2}\sqrt{\sum_{j\in[\![k]\!]} x_j^4} \vee a^2_k\right\}.$$

Additionally, we define

$$(r^{\mathrm{d}}_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(x_\bullet))^2 := \min_{k\in\mathcal{K}}\left\{\mathrm{v}_k^{-2}\mathrm{m}_k(x^2_\bullet) \vee a^2_k\right\} = \min_{k\in\mathcal{K}}\left\{\mathrm{v}_k^{-2}\max_{j\in[\![k]\!]}(x^2_j) \vee a^2_k\right\}. \tag{1.5.13}$$

We first present an adaptive upper bound in terms of the reparametrised noise level $(\varepsilon^\circ_\bullet)^2 = \varepsilon^2_\bullet + (\theta^\circ_\bullet)^2\sigma^2_\bullet$ and the factor $\delta_{\mathcal{K}} := (1 \vee \log|\mathcal{K}|)^{1/4}$. Again, the upper bound has two regimes, which determine whether we obtain an adaptive factor $\delta_{\mathcal{K}}$ or $\delta^2_{\mathcal{K}}$.

Proposition 1.5.6 (Adaptive upper bound – direct max-test). For $\alpha \in (0,1)$ define $\overline{A}_\alpha^2 := \mathrm{R}^2 + \mathrm{c}\left(5L_{\alpha/2} + 15L_{\alpha/2}^2 + 5\right)$. Then for all $A \geq \overline{A}_\alpha$ we obtain

$$\sup_{(a_\bullet,\mathrm{v}_\bullet)\in\mathcal{A}\times\mathcal{V}} \mathcal{R}\left(\Delta^{\mathrm{d}}_{\mathcal{K},\alpha/2} \mid \Theta^{\mathrm{R}}_{a_\bullet}, \Lambda^{\mathrm{c}}_{\mathrm{v}_\bullet}, \theta^\circ_\bullet, A\left\{\rho^{\mathrm{d}}_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\mathcal{K}}\varepsilon^\circ_\bullet) \vee r^{\mathrm{d}}_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta^2_{\mathcal{K}}\varepsilon^\circ_\bullet)\right\}\right) \leq \alpha.$$

Proof of Proposition 1.5.6. The proof follows along the lines of the proof of Proposition 1.5.1 using Proposition 1.4.1 rather than Proposition 1.2.1 and we omit the details. □

Remark 1.5.7 (Adaptive factor $\delta^2_{\mathcal{K}}$ vs. $\delta_{\mathcal{K}}$). *The upper bound in Proposition 1.5.6 consists of two terms similar to the upper bound in Proposition 1.5.1. In contrast to $r_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta^2_{\mathcal{K}}\varepsilon^\circ_\bullet)$ in Proposition 1.5.1 the term $r^{\mathrm{d}}_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta^2_{\mathcal{K}}\varepsilon^\circ_\bullet)$ in Proposition 1.5.6 is generally not negligible compared to $\rho^{\mathrm{d}}_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\mathcal{K}}\varepsilon^\circ_\bullet)$ if the effective noise level $(\theta^\circ_\bullet)^2\sigma^2_\bullet$ determines the radius. That is, e.g. in the homoscedastic case $r^{\mathrm{d}}_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta^2_{\mathcal{K}}\sigma_\bullet)$ and $\rho^{\mathrm{d}}_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta^2_{\mathcal{K}}\sigma_\bullet)$ are of the same order. Hence, in this case, we obtain an adaptive factor $\delta^2_{\mathcal{K}}$. If, however, the noise level $\varepsilon_\bullet$ governs the radius, $r^{\mathrm{d}}_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta^2_{\mathcal{K}}\varepsilon_\bullet)$ is often negligible compared with $\rho^{\mathrm{d}}_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\mathcal{K}}\varepsilon^\circ_\bullet)$, yielding an adaptive factor $\delta_{\mathcal{K}}$. Again, we give sufficient conditions for this to happen. Similar to Remark 1.5.2 consider $k^{\mathrm{d}}_\star := k^{\mathrm{d}}_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\mathcal{K}}\varepsilon_\bullet)$. Assume there exists a $C > 0$ such that (compare assumption (1.5.9))*

$$\sqrt{k^{\mathrm{d}}_\star}\mathrm{m}_{k^{\mathrm{d}}_\star}(\varepsilon^2_\bullet) \leq C\mathrm{q}_{k^{\mathrm{d}}_\star}(\varepsilon^2_\bullet), \tag{1.5.14}$$

and assume additionally that

$$\delta^2_{\mathcal{K}} \leq C\sqrt{k^{\mathrm{d}}_\star}. \tag{1.5.15}$$

Then, trivially,

$$\mathrm{m}_{k^{\mathrm{d}}_\star}\left(\delta^4_{\mathcal{K}}\varepsilon^2_\bullet\right) \leq C\sqrt{k^{\mathrm{d}}_\star}\mathrm{m}_{k^{\mathrm{d}}_\star}\left(\delta^2_{\mathcal{K}}\varepsilon^2_\bullet\right) \leq C^2\mathrm{q}_k\left(\delta^2_{\mathcal{K}}\varepsilon^2_\bullet\right),$$

and, hence, $r^{\mathrm{d}}_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta^2_{\mathcal{K}}\varepsilon_\bullet) \leq C\rho^{\mathrm{d}}_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\mathcal{K}}\varepsilon_\bullet)$. □

Next, we want to formulate the upper bound in Proposition 1.5.6 in terms of the noise levels $\varepsilon_\bullet$ and $\theta^\circ_\bullet\sigma_\bullet$. Similar to the previous section, we choose collections $\mathcal{K}_{\varepsilon_\bullet}$ and $\mathcal{K}_{\sigma_\bullet}$ and set $\mathcal{K} := \mathcal{K}_{\varepsilon_\bullet} \cap \mathcal{K}_{\sigma_\bullet}$ with $\delta_{\varepsilon_\bullet} := \delta_{\mathcal{K}_{\varepsilon_\bullet}}$, $\delta_{\sigma_\bullet} := \delta_{\mathcal{K}_{\sigma_\bullet}}$ and, hence, $\delta_{\mathcal{K}} \leq \delta_{\varepsilon_\bullet} \wedge \delta_{\sigma_\bullet}$. The assumption (1.5.16) states that the remainder term evaluated in $\delta^2_{\varepsilon_\bullet}\varepsilon_\bullet$ is negligible. As discussed in Remark 1.5.7 the remainder term evaluated in $\delta^2_{\sigma_\bullet}\theta^\circ_\bullet\sigma_\bullet$ is generally not negligible.

Theorem 1.5.8 (Adaptive upper bound - direct max-test). Let $\mathcal{K} := \mathcal{K}_{\varepsilon_\bullet} \cap \mathcal{K}_{\sigma_\bullet}$, $\delta_{\varepsilon_\bullet} := \delta_{\mathcal{K}_{\varepsilon_\bullet}}$ and $\delta_{\sigma_\bullet} := \delta_{\mathcal{K}_{\sigma_\bullet}}$. Assume there exists a $C \geq 1$ such that

$$r^{\mathrm{d}}_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta^2_{\varepsilon}\varepsilon_\bullet) \leq C\rho^{\mathrm{d}}_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\varepsilon_\bullet}\varepsilon_\bullet) \tag{1.5.16}$$

for $a_\bullet \in \mathcal{A}$ and $\mathrm{v}_\bullet \in \mathcal{V}$. Then, for each $\alpha \in (0,1)$ with $\overline{A}^2_\alpha := 2\left(\mathrm{R}^2 + \mathrm{c}\left(5L_{\alpha/2} + 15L^2_{\alpha/2} + 5\right)\right)$ it follows for all $A \geq \overline{A}_\alpha$

$$\sup_{a_\bullet\in\mathcal{A}} \mathcal{R}\left(\Delta_{\mathcal{K},\alpha/2} \mid \Theta^{\mathrm{R}}_{a_\bullet}, \Lambda^{\mathrm{c}}_{\mathrm{v}_\bullet}, \theta^\circ_\bullet, AC\left\{\rho^{\mathrm{d}}_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\varepsilon_\bullet}\varepsilon_\bullet) \vee \rho^{\mathrm{d}}_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta^2_{\sigma_\bullet}\theta^\circ_\bullet\sigma_\bullet)\right\}\right) \leq \alpha.$$

Proof of Theorem 1.5.8. We have the elementary inequalities

$$2\left\{\rho^{\mathrm{d}}_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\varepsilon_\bullet}\varepsilon_\bullet) \vee \rho^{\mathrm{d}}_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\sigma_\bullet}\theta^\circ_\bullet\sigma_\bullet)\right\} \geq \rho^{\mathrm{d}}_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\mathcal{K}}\varepsilon^\circ_\bullet),$$

$$2\left\{r^{\mathrm{d}}_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta^2_{\varepsilon_\bullet}\varepsilon_\bullet) \vee r^{\mathrm{d}}_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta^2_{\sigma_\bullet}\theta^\circ_\bullet\sigma_\bullet)\right\} \geq r^{\mathrm{d}}_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta^2_{\mathcal{K}}\varepsilon^\circ_\bullet)$$

and

$$\rho^{\mathrm{d}}_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta^2_{\sigma_\bullet}\theta^\circ_\bullet\sigma_\bullet) \geq r^{\mathrm{d}}_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta^2_{\sigma_\bullet}\theta^\circ_\bullet\sigma_\bullet)$$

The assertion is now an immediate consequence of Proposition 1.5.6 and assumption (1.5.16). □

Remark 1.5.9 (Optimality w.r.t $\varepsilon_\bullet$, suboptimality w.r.t $\sigma_\bullet$). *Comparing the upper bounds in Theorem 1.5.3 and Theorem 1.5.8 for the indirect and the direct* max*-tests shows that there appears an additional factor $\delta_{\sigma_\bullet}$ in the term $\rho^{\mathrm{d}}_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta^2_{\sigma_\bullet}\theta^\circ_\bullet\sigma_\bullet)$. However, the radius $\left(\rho^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\theta^\circ_\bullet\sigma_\bullet)\right)^2$ achieved by the direct test is generally already much larger than the radius $\rho^2_{a_\bullet,\mathrm{v}_\bullet}(\theta^\circ_\bullet\sigma_\bullet)$ achieved by the indirect test and the additional deterioration by a factor $\delta_{\sigma_\bullet}$ is negligible compared with it. In other word, if the $\sigma_\bullet$-part of the radius determines its behaviour, the direct test already performs suboptimally and the additional factor plays an inconsequential role. On the other hand, if the $\varepsilon_\bullet$-part of the radius determines its behaviour, then $\rho^{\mathrm{d}}_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\varepsilon_\bullet}\varepsilon_\bullet)$ is typically of the same order as $\rho_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\varepsilon_\bullet}\varepsilon_\bullet)$ and, hence, optimal.* □

Illustration 1.5.10 (Homoscedastic case). Consider the homoscedastic case and the smoothness and ill-posedness assumptions introduced in Illustration 1.2.5. Choosing a geometric grid $\mathcal{K} := \mathcal{K}_g := \{2^j : j \in [\![4\,|\log_2\varepsilon|\,]\!] \cap [\![4\,|\log_2\sigma|\,]\!]\} \cup \{1\}$ with $\delta_{\mathcal{K}} \leq \delta_{\varepsilon_\bullet} \wedge \delta_{\sigma_\bullet}$, $\delta_{\varepsilon_\bullet} \sim (\log|\log\varepsilon|)^{1/4}$, $\delta_{\sigma_\bullet} \sim (\log|\log\sigma|)^{1/4}$ yields

$$\rho^{\mathrm{d}}_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\varepsilon_\bullet}\varepsilon_\bullet) \vee \rho^{\mathrm{d}}_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta^2_{\sigma_\bullet}\theta^\circ_\bullet\sigma_\bullet) \leq C\left\{\rho^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\delta_{\varepsilon_\bullet}\varepsilon_\bullet) \vee \rho^{\mathrm{d}}_{a_\bullet,\mathrm{v}_\bullet}(\delta^2_{\sigma_\bullet}\theta^\circ_\bullet\sigma_\bullet)\right\}$$

for a constant $C \geq 1$ chosen uniformly for $s \in [s_\star, s^\star]$ and $p \in [p_\star, p^\star]$. That is, the minimisation over $\mathcal{K}$ approximates the minimisation over $\mathbb{N}$ sufficiently well.
Moreover, for **mildly ill-posed** models we have

$$r^{\mathrm{d}}_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta^2_{\mathcal{K}}\varepsilon_\bullet) \leq C\rho^{\mathrm{d}}_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\varepsilon_\bullet}\varepsilon_\bullet)$$

for some $C \geq 1$ uniformly for all $s \in [s_\star, s^\star]$ and $p \in [p_\star, p^\star]$, since the conditions (1.5.14) and (1.5.15) are fulfilled.
In a **severely ill-posed case** the terms $\rho^{\mathrm{d}}_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\mathcal{K}}\varepsilon^\circ_\bullet)$, $r^{\mathrm{d}}_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta^2_{\mathcal{K}}\varepsilon^\circ_\bullet)$ and $\rho^{\mathrm{d}}_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\varepsilon^\circ_\bullet)$ are all of the same order and the adaptive factor has no effect on the rate. We present the resulting adaptive radii from Theorem 1.5.8 in the table below.

Order of the adaptive radius $\left(\rho^{\mathrm{d}}_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\varepsilon_\bullet}\varepsilon_\bullet) \vee \rho^{\mathrm{d}}_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta^2_{\sigma_\bullet}\theta^\circ_\bullet\sigma_\bullet)\right)^2$ for the geometric grid $\mathcal{K} := \mathcal{K}_g := \{2^j : j \in [\![4\,|\log_2\varepsilon|\,]\!] \cap [\![4\,|\log_2\sigma|\,]\!]\} \cup \{1\}$ in the homoscedastic case $\varepsilon_\bullet = (\varepsilon)_{j\in\mathbb{N}}$ and $\sigma_\bullet = (\sigma)_{j\in\mathbb{N}}$

a_j (smooth.)	v_j (ill-posed.)	$(\rho^{\mathrm{d}}_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\varepsilon_\bullet}\varepsilon_\bullet))^2$, $\theta^\circ_\bullet \in \ell^2$	$(\rho^{\mathrm{d}}_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta^2_{\sigma_\bullet}\theta^\circ_\bullet\sigma_\bullet))^2$, $\theta^\circ_\bullet = (j^{-t})_{j\in\mathbb{N}}$
j^{-s}	j^{-p}	$\left((\log\|\log\varepsilon\|)^{\frac{1}{4}}\varepsilon\right)^{\frac{8s}{4s+4p+1}}$	$((\log\|\log\sigma\|)^{\frac{1}{2}}\sigma)^{\frac{2s}{s+p}}$
j^{-s}	e^{-j^p}	$\|\log\varepsilon\|^{-\frac{2s}{p}}$	$\|\log\sigma\|^{-\frac{2s}{p}}$
e^{-j^s}	j^{-p}	$\varepsilon^2(\log\|\log\varepsilon\|)^{\frac{1}{2}}\,\|\log\varepsilon\|^{\frac{4p+1}{2s}}$	$\sigma^2(\log\|\log\sigma\|)\,\|\log\sigma\|^{\frac{2p}{s}}$

We shall stress that the orders of the upper bounds in terms of $\varepsilon_\bullet = (\varepsilon)_{j\in\mathbb{N}}$, i.e. $\rho^{\mathrm{d}}_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\varepsilon_\bullet}\varepsilon_\bullet)$ (direct test) and $\rho_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\varepsilon_\bullet}\varepsilon_\bullet)$ (indirect test), coincide in all three cases. Therefore, the direct test performs optimally with an adaptive factor $\delta_{\varepsilon_\bullet}$ if the $\varepsilon_\bullet$-terms govern the radii, this e.g. happens in the case $\varepsilon \geq \sigma$. However, the upper bound $\rho^{\mathrm{d}}_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta^2_{\sigma_\bullet}\sigma_\bullet)$ in terms of $\sigma_\bullet$ obtained for the direct test is generally much slower than the optimal radius $\rho_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\sigma_\bullet}\sigma_\bullet)$ achieved by the indirect test.

As in Illustration 1.5.5 in case of super smoothness $a_\bullet = (e^{-j^s})_{j\in\mathbb{N}}$ we consider a smaller geometric grid for adaptation to $s_\star > 0$

$$\mathcal{K} := \mathcal{K}_{s_\star} := \left\{2^j : j \in [\![\frac{1}{2s_\star}\log_2|\log\varepsilon|\,]\!] \cap [\![\frac{1}{2s_\star}\log_2|\log\sigma|\,]\!]\right\} \cup \{1\}$$

and an adaptive factor $\delta_{\mathcal{K}_{s_\star}} \leq \delta_\varepsilon \wedge \delta_\sigma$ with $\delta_\varepsilon \sim (\log\log|\log\varepsilon|)^{\frac{1}{4}}$ and $\delta_\sigma \sim (\log\log|\log\sigma|)^{\frac{1}{4}}$.

Order of the adaptive radius $\left(\rho^{\mathrm{d}}_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\varepsilon_\bullet}\varepsilon_\bullet) \vee \rho^{\mathrm{d}}_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta^2_{\sigma_\bullet}\theta^\circ_\bullet\sigma_\bullet)\right)^2$ for the geometric grid $\mathcal{K} := \mathcal{K}_{s_\star} := \left\{2^j : j \in [\![\frac{1}{2s_\star}\log_2|\log\varepsilon|\,]\!] \cap [\![\frac{1}{2s_\star}\log_2|\log\sigma|\,]\!]\right\}$ in the homoscedastic case $\varepsilon_\bullet = (\varepsilon)_{j\in\mathbb{N}}$ and $\sigma_\bullet = (\sigma)_{j\in\mathbb{N}}$

a_j (smooth.)	v_j (ill-posed.)	$(\rho^{\mathrm{d}}_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta_{\varepsilon_\bullet}\varepsilon_\bullet))^2$ $\theta^\circ_\bullet \in \ell^2$	$(\rho^{\mathrm{d}}_{\mathcal{K},a_\bullet,\mathrm{v}_\bullet}(\delta^2_{\sigma_\bullet}\theta^\circ_\bullet\sigma_\bullet))^2$ $\theta^\circ_\bullet = (j^{-t})_{j\in\mathbb{N}}$
e^{-j^s}	j^{-p}	$\varepsilon^2(\log\log|\log\varepsilon|)^{\frac{1}{2}}\,|\log\varepsilon|^{\frac{4p+1}{2s}}$	$\sigma^2(\log\log|\log\sigma|)\,|\log\sigma|^{\frac{2p}{s}}$

1.6 Adaptive lower bound

In this section we provide conditions under which a deterioration of the minimax testing radius is unavoidable for adaptation over $\left\{\Theta^{\mathrm{R}}_{a_\bullet} : a_\bullet \in \mathcal{A}\right\}$, where $\mathcal{A} \subseteq \ell^2$ is a class of regularity sequences.

Proposition 1.6.1 (Adaptive lower bound). Let $\alpha \in (0,1)$, $\delta = \delta_{\varepsilon_\bullet} \geq 1$ and let $\mathrm{v}_\bullet \in \mathcal{V}$ be fixed. Assume that there exists a collection of N regularity sequences $\{a^j_\bullet : j \in [\![N]\!]\} \subseteq \mathcal{A}$, where we abbreviate for $j \in [\![N]\!]$

$$\rho_j := \rho_{a^j_\bullet,\mathrm{v}_\bullet}(\delta\varepsilon_\bullet) \quad \text{with associated dimension parameters} \quad k_j := k_{a^j_\bullet,\mathrm{v}_\bullet}(\delta\varepsilon_\bullet)$$

such that the following four conditions are satisfied.

(C1) The collection is ordered such that $k_l \leq k_m$ and $\delta^2\rho_l \leq \rho_m$, whenever $l < m$.

(C2) There exists a finite constant $c_\alpha > 0$ such that $\exp(c_\alpha\delta^4) \leq N\alpha^2$.

(C3) There exists a constant $\eta \in (0,1]$ such that

$$\eta \leq \min_{j\in[\![N]\!]} \frac{(a^j_{k_j})^2 \wedge \mathrm{q}_{k_j}\left(\frac{(\delta\varepsilon_\bullet)^2}{\mathrm{v}^2_\bullet}\right)}{(a^j_{k_j})^2 \vee \mathrm{q}_{k_j}\left(\frac{(\delta\varepsilon_\bullet)^2}{\mathrm{v}^2_\bullet}\right)} = \min_{j\in[\![N]\!]} \frac{(a^j_{k_j})^2 \wedge \mathrm{q}_{k_j}\left(\frac{(\delta\varepsilon_\bullet)^2}{\mathrm{v}^2_\bullet}\right)}{\rho^2_j}.$$

Then, with

$$\underline{A}^2_\alpha := \eta\left(\mathrm{R}^2 \wedge \sqrt{\log(1+\alpha^2)} \wedge \sqrt{c_\alpha}\right)$$

we obtain for all $A \in [0, \underline{A}_\alpha]$

$$\inf_\Delta \sup_{a_\bullet \in \mathcal{A}} \mathcal{R}\left(\Delta \mid \Theta^{\mathrm{R}}_{a_\bullet}, \Lambda^{\mathrm{c}}_{\mathrm{v}_\bullet}, \theta^\circ_\bullet, A\rho_{a_\bullet,\mathrm{v}_\bullet}(\delta\varepsilon_\bullet)\right) \geq 1-\alpha.$$

Proof of Proposition 1.6.1. The proof generalises the reduction scheme of Proposition 1.3.1 to multiple classes of alternatives.

Reduction step. We write $\mathbb{P}_0 := \mathbb{P}^{\varepsilon_\bullet,\sigma_\bullet}_{\theta^\circ_\bullet,\mathrm{v}_\bullet}$. Let $\mathbb{P}_{1,m}$ be a mixing measure over the $\underline{A}_\alpha\rho_m$-separated alternative $\Theta^{\mathrm{R}}_{a^m_\bullet}$ and consider the uniform mixture $\mathbb{P}_1 := \frac{1}{N}\sum_{m\in[\![N]\!]}\mathbb{P}_{1,m}$ over all $m \in [\![N]\!]$. Replacing the supremum over all $a_\bullet \in \mathcal{A}$ with a maximum over all $a^m_\bullet, m \in [\![N]\!]$ and then the maximum by the average over $m \in [\![N]\!]$ and combining this with the reduction step of Proposition 1.3.1 it is easily seen that (cp. the proof of Proposition 4.6.1 for more details)

$$\inf_\Delta \sup_{a_\bullet \in \mathcal{A}} \mathcal{R}\left(\Delta \mid \Theta^{\mathrm{R}}_{a_\bullet}, \Lambda^{\mathrm{c}}_{\mathrm{v}_\bullet}, \theta^\circ_\bullet, A\rho_{a_\bullet,\mathrm{v}_\bullet}(\delta\varepsilon_\bullet)\right) \geq 1-\sqrt{\frac{\chi^2(\mathbb{P}_0,\mathbb{P}_1)}{2}}.$$

Definition of the mixtures. For each $m \in [\![N]\!]$ we introduce deviations from the null $\tilde{\theta}^m_\bullet \in \ell^2$ by setting

$$\tilde{\theta}^m_j := \begin{cases} \frac{\underline{A}_\alpha\rho_m}{\mathrm{q}_{k_m}(\frac{\delta^2\varepsilon^2_\bullet}{\mathrm{v}^2_\bullet})}\frac{\delta^2\varepsilon^2_j}{\mathrm{v}^2_j} & \text{for } j \in [\![k_m]\!], \\ 0 & \text{otherwise.}\end{cases}$$

Note that $\tilde{\theta}^m_\bullet \in \Theta^{\mathrm{R}}_{a^j_\bullet} \cap \ell^2_{\underline{A}_\alpha\rho_m}$ follows exactly as the in proof of Proposition 1.3.1. Thus, we define $\mathbb{P}_{1,m} := \frac{1}{2^{k_m}}\sum_{\tau\in\{\pm\}^{k_m}}\mathbb{P}^{\varepsilon_\bullet,\sigma_\bullet}_{\theta^\circ_\bullet+\tilde{\theta}^{m,\tau}_\bullet,\mathrm{v}_\bullet}$ where $\tilde{\theta}^{m,\tau}_j = \tau_j\tilde{\theta}^m_j\mathbb{1}_{\{j\in[\![k_m]\!]\}}$.

Bound for the χ^2-divergence. Arguing as in the proof of Proposition 1.3.1 and applying Lemma A.3.1 from the appendix yields

$$\chi^2(\mathbb{P}_0,\mathbb{P}_1) \leq \frac{1}{N^2}\sum_{m,l\in[\![N]\!]}\exp\left(\frac{1}{2}\mathrm{q}^2_{k_m\wedge k_l}(\nu^2_\bullet\tilde{\theta}^m_\bullet\tilde{\theta}^l_\bullet/\varepsilon^2_\bullet)\right)-1.$$

We insert the definition of the deviations $\tilde{\theta}^m_\bullet$, $\tilde{\theta}^l_\bullet$, exploit conditions (C1) and (C3) and obtain for $l \leq m$

$$\mathrm{q}^2_{k_m\wedge k_l}(\nu^2_\bullet\tilde{\theta}^m_\bullet\tilde{\theta}^l_\bullet/\varepsilon^2_\bullet) \leq 2\xi^2\delta^4\frac{\rho^2_l}{\rho^2_m}.$$

with $\xi = \mathrm{R}^2 \wedge \sqrt{\log(1+\alpha^2)} \wedge \sqrt{c_\alpha}$. Hence, by splitting the sum into two parts ($m = l$ and $m \neq l$) we get

$$\chi^2(\mathbb{P}_0,\mathbb{P}_1) \leq \frac{1}{N}\exp(c_\alpha\delta^4) + \frac{N(N-1)}{N^2}\exp(\log(1+\alpha^2)) - 1 \leq 2\alpha^2,$$

where we used both (C1) (first inequality) and (C2) (second inequality). Inserting this bound into the reduction step completes the proof. ☐

Remark 1.6.2 (Conditions of Proposition 1.6.1). *Let us comment on the conditions of Proposition 1.6.1. Condition (C1) requires $\mathcal{A}$ to contain distinguishable elements $a^m_\bullet$, which result in significantly different radii ρ_m. This is a sensible condition: assume all elements in $\mathcal{A}$ yield the same separation radius and the same optimal dimension. Naturally, adaptation can then be achieved without a loss. We only expect to pay for adaptivity if we need to incorporate various dimension parameters k in our adaptation procedure. Condition (C2) gives an upper bound for the maximal size of the adaptive factor and (C3) is a balancing condition, which already appears in the non-adaptive lower bound Proposition 1.3.1, but now has to hold uniformly over the collection.* ☐

Proposition 1.6.1 gives general conditions on the collection $\left\{\Theta^{\mathrm{R}}_{a_\bullet} : a_\bullet \in \mathcal{A}\right\}$ of regularity alternatives, which make adaptation without a loss impossible. Next, we demonstrate how to use it in the homoscedastic setting to show that the adaptive factors that we obtain in Illustration 1.5.5 are minimal.

Theorem 1.6.3 (Minimal adaptive factor, polynomial decay). We consider the homoscedastic case $\varepsilon_\bullet = (\varepsilon)_{j\in\mathbb{N}}$. Let $\mathcal{A}$ be non-trivial with respect to polynomial decay for some $s^\star > s_\star > \frac{1}{2}$, i.e.

$$\{(j^{-s})_{j\in\mathbb{N}} : s \in [s_\star, s^\star]\} \subseteq \mathcal{A}$$

and let $\mathrm{v}_\bullet := (j^{-p})_{j\in\mathbb{N}}$, $p > 0$ be fixed.
For $\alpha \in (0,1)$ set $\underline{A}^2_\alpha := \eta\left(\mathrm{R}^2 \wedge \sqrt{\log(1+\alpha^2)} \wedge 1/2\right)$ with η as in Proposition 1.6.1. There exists an $\tilde{\varepsilon} \in (0,1)$ such that for all $0 < \varepsilon < \tilde{\varepsilon}$ and all $A \le \underline{A}_\alpha$ we have

$$\inf_{\Delta} \sup_{a_\bullet\in\mathcal{A}} \mathcal{R}\left(\Delta \mid \Theta^{\mathrm{R}}_{a_\bullet}, \Lambda^{\mathrm{c}}_{\mathrm{v}_\bullet}, \theta^\circ_\bullet, A\rho_{a_\bullet,\mathrm{v}_\bullet}(\delta\varepsilon_\bullet)\right) \ge 1-\alpha$$

with $\delta = (\log|\log\varepsilon|)^{\frac{1}{4}}$, i.e. δ is a lower bound for the minimal adaptive factor over $\mathcal{A}$.

Proof of Theorem 1.6.3. We construct a collection $\mathcal{A}_N := \{a^m_\bullet \in \mathcal{A} : m \in [\![N]\!]\} \subseteq \mathcal{A}$ such that (C1) – (C3) of Proposition 1.6.1 are satisfied.
Definition of the collection. We have seen in Illustration 1.2.5 that the minimax radius in case of ordinary smoothness and mild ill-posedness is of order $\rho^2_{a_\bullet,\mathrm{v}_\bullet}(\delta\varepsilon_\bullet) \sim (\delta\varepsilon)^{e(s)}$ with the exponent $e(s) := \frac{8s}{4s+4p+1}$. The exponent $e(s) = \frac{8s}{4s+4p+1} = 2 - \frac{8p+2}{4s+4p+1}$ is monotonically increasing in s, hence the corresponding regularity parameters result in radii with exponents in the interval $[e(s_\star), e(s^\star)] =: [e_\star, e^\star]$. A grid of size N on $[e_\star, e^\star]$ induces a grid on $[s_\star, s^\star]$, which in turn defines a grid on $\mathcal{A}$. Let $d := \frac{e^\star - e_\star}{N}$ and

$$\mathcal{G}_s := \{s_m : e(s_m) = e^\star - md, m \in \{0,\dots,N-1\}\},$$

which we use to define our collection of regularity sequences

$$\mathcal{G}_{a_\bullet} := \{(j^{-s})_{j\in\mathbb{N}} : s \in \mathcal{G}_s\}.$$

Tedious, but elementary calculations (comparable to those in the proof of Theorem 4.6.3 and thus omitted) show that (C1) – (C3) are satisfied with $N = \lfloor \frac{e^\star - e^\star}{4} \frac{|\log(\delta\varepsilon)|}{\log\delta} \rfloor$, $\delta = (\log|\log\varepsilon|)^{1/4}$ and ε small enough. ☐

Theorem 1.6.4 (Minimal adaptive factor, exponential decay). We consider the homoscedastic case $\varepsilon_\bullet = (\varepsilon)_{j\in\mathbb{N}}$. Let $\mathcal{A}$ be non-trivial with respect to exponential decay for some $s^\star > s_\star > 0$, i.e.

$$\left\{(e^{-j^s})_{j\in\mathbb{N}} : s \in [s_\star, s^\star]\right\} \subseteq \mathcal{A}$$

and let $\mathrm{v}_\bullet := (j^{-p})_{j\in\mathbb{N}}$, $p > 0$ be fixed.
For $\alpha \in (0,1)$ set $\underline{A}^2_\alpha := \eta\left(\mathrm{R}^2 \wedge \sqrt{\log(1+\alpha^2)} \wedge 1/2\right)$ with η as in Proposition 1.6.1. There exists an $\tilde{\varepsilon} \in (0,1)$ such that for all $0 < \varepsilon < \tilde{\varepsilon}$ and all $A \le \underline{A}_\alpha$ we have

$$\inf_{\Delta} \sup_{a_\bullet\in\mathcal{A}} \mathcal{R}\left(\Delta \mid \Theta^{\mathrm{R}}_{a_\bullet}, \Lambda^{\mathrm{c}}_{\mathrm{v}_\bullet}, \theta^\circ_\bullet, A\rho_{a_\bullet,\mathrm{v}_\bullet}(\delta\varepsilon_\bullet)\right) \ge 1-\alpha$$

with $\delta = (\log\log|\log\varepsilon|)^{\frac{1}{4}}$, i.e. δ is a lower bound for the minimal adaptive factor over $\mathcal{A}$.

Proof of Theorem 1.6.3. We construct a collection $\mathcal{A}_N := \{a_\bullet^m \in \mathcal{A} : m \in [\![N]\!]\} \subseteq \mathcal{A}$ such that (C1) – (C3) of Proposition 1.6.1 are satisfied.

Definition of the collection. We have seen in Illustration 1.2.5 that the minimax radius in case of super smoothness and mild ill-posedness is of order $\rho^2_{a_\bullet, \mathrm{v}_\bullet}(\delta\varepsilon_\bullet) \sim (\delta\varepsilon)^2(\log\delta\varepsilon)^{-e(s)}$ with the exponent $e(s) := \frac{4p+1}{2s}$. The exponent is monotonically decreasing in s, hence the corresponding regularity parameters result in radii with exponents in the interval $[e(s^\star), e(s_\star)] =: [e_\star, e^\star]$. A grid of size N on $[e_\star, e^\star]$ induces a grid on $[s_\star, s^\star]$, which in turn defines a grid on $\mathcal{A}$. Let $d := \frac{e^\star - e_\star}{N}$ and

$$\mathcal{G}_s := \{s_m : e(s_m) = e_\star + md, m \in \{0, \dots, N-1\}\},$$

which we use to define our collection of regularity sequences

$$\mathcal{G}_{a_\bullet} := \left\{(e^{-j^s})_{j\in\mathbb{N}} : s \in \mathcal{G}_s\right\}.$$

Again, the calculations to verify (C1) – (C3) are elementary but tedious (and omitted since they are comparable to those in the proof of Theorem 4.6.4). With $N = \lfloor \frac{e^\star - e^\star}{4} \frac{\log|\log(\delta\varepsilon)|}{\log\delta} \rfloor$, $\delta = (\log\log|\log\varepsilon|)^{1/4}$ and ε small enough the assertion follows. $\square$

Summarizing, Theorem 1.6.3 and Theorem 1.6.4 establish the optimality of the adaptive factors with respect to the noise level ε obtained in Illustration 1.5.5 and Illustration 1.5.10. That is, we have shown that for adaptation in an **ordinary smooth – mildly ill-posed** model, the minimal adaptive factor is given by $(\log|\log\varepsilon|)^{1/4}$. Moreover, in a **super smooth – mildly ill-posed** model the cost to pay for adaptation is only of order $(\log\log|\log\varepsilon|)^{1/4}$ and it is unavoidable. We point out that in the third case (**ordinary smooth – severely ill-posed**) the rates are very slow (i.e. logarithmic in the noise level) due to the severe ill-posedness. Moreover, the optimal dimension does not depend on the smoothness parameter and, hence, the indirect and the direct testing procedure are automatically adaptive with respect to smoothness. However, even if we carry out our adaptation procedure for the direct testing procedure to make it adaptive with respect to the ill-posedness of the model, the additional factor (caused by the aggregation) does not have an effect on the rate.

Appendix A

Auxiliary results

A.1 Non-central χ^2-random variables

Lemma A.1.1 (Quantiles of (non-central) χ^2-random variables). For $\mu_\bullet \in \ell^2$ and $e_\bullet \in \mathbb{R}_+^{\mathbb{N}}$ let $Z_\bullet \sim \mathbb{P}^{e_\bullet}_{\mu_\bullet} := \mathcal{N}(\mu_\bullet, e_\bullet^2)$. For each $k \in \mathbb{N}$ define $\mathrm{Q}_k := \sum_{j\in[\![k]\!]} Z_j^2$ and denote by $\mathbb{Q}^{e_\bullet}_{\mu_\bullet,k}$ its distribution, i.e. $\mathrm{Q}_k \sim \mathbb{Q}^{e_\bullet}_{\mu_\bullet,k}$ and by $\mathrm{q}^{e_\bullet}_{\mu_\bullet,k}(u)$ the $(1-u)$-quantiles of $\mathrm{Q}^{e_\bullet}_{\mu_\bullet,k}$, i.e. $\mathbb{P}^{e_\bullet}_{\mu_\bullet}\left(\mathrm{Q}_k \le \mathrm{q}^{e_\bullet}_{\mu_\bullet,k}(u)\right) = 1-u$. For any $k \in \mathbb{N}$ and $u \in (0,1)$ with $L_u := \sqrt{|\log u|}$ we have

$$\mathrm{q}^{e_\bullet}_{0_\bullet,k}(u) \le \mathrm{q}_k^2(e_\bullet) + 2L_u \mathrm{q}_k(e_\bullet^2) + 2L_u^2 \mathrm{m}_k(e_\bullet^2) \tag{A.1.1}$$

$$\le \mathrm{q}_k^2(e_\bullet) + 2(L_u + L_u^2)\mathrm{q}_k(e_\bullet^2), \tag{A.1.2}$$

$$\mathrm{q}^{e_\bullet}_{\mu_\bullet,k}(1-u) \ge \mathrm{q}_k^2(e_\bullet) + \frac{4}{5}\mathrm{q}_k^2(\mu_\bullet) - 2\left(L_u + 5L_u^2\right)\mathrm{q}_k(e_\bullet^2). \tag{A.1.3}$$

Proof of Lemma A.1.1. We start our proof with the observations that

$$\mathbb{E}^{e_\bullet}_{\mu_\bullet}\mathrm{Q}_k = \sum_{j\in[\![k]\!]}(e_j^2+\mu_j^2) = \mathrm{q}_k^2(e_\bullet) + \mathrm{q}_k^2(\mu_\bullet)$$

$$\Sigma_k := \frac{1}{2}\sum_{j\in[\![k]\!]}\mathrm{var}^{e_\bullet}_{\mu_\bullet}(Z_j^2) = \sum_{j\in[\![k]\!]} e_j^2(e_j^2+2\mu_j^2) = \mathrm{q}_k^2(e_\bullet^2) + 2\mathrm{q}_k^2(\mu_\bullet e_\bullet)$$

since $\mathrm{var}^{e_\bullet}_{\mu_\bullet}(Z_j^2) = \mathbb{E}^{e_\bullet}_{\mu_\bullet}(Z_j^4) - \left(\mathbb{E}^{e_\bullet}_{\mu_\bullet}(Z_j^2)\right)^2 = \mu_j^4 + 6\mu_j^2e_j^2 + 3e_j^4 - (\mu_j^2+e_j^2)^2 = 4\mu_j^2e_j^2 + 2e_j^4$. Moreover, we have

$$\sqrt{\mathrm{q}_k^2(e_\bullet^2)} = \mathrm{q}_k^2(e_\bullet^2) \ge \mathrm{m}_k(e_\bullet^2),$$

which we use below without further reference. Due to Birgé [2001] (Lemma 8.1) it follows for all $x > 0$

$$\mathbb{P}^{e_\bullet}_{\mu_\bullet}\left(\mathrm{Q}_k - \mathbb{E}^{e_\bullet}_{\mu_\bullet}(\mathrm{Q}_k) \ge 2\sqrt{\Sigma_k x} + 2\mathrm{m}_k(e_\bullet^2)x\right) \le \exp(-x)$$

$$\mathbb{P}^{e_\bullet}_{\mu_\bullet}\left(\mathrm{Q}_k - \mathbb{E}^{e_\bullet}_{\mu_\bullet}(\mathrm{Q}_k) \le -2\sqrt{\Sigma_k x}\right) \le \exp(-x),$$

which for all $u \in (0,1)$ with $L_u = \sqrt{|\log u|}$ implies

$$\mathrm{q}^{e_\bullet}_{\mu_\bullet,k}(u) \le \mathrm{q}_k^2(e_\bullet) + \mathrm{q}_k^2(\mu_\bullet) + 2\sqrt{L_u^2\Sigma_k} + 2L_u^2\mathrm{m}_k(e_\bullet^2),$$

$$\mathrm{q}^{e_\bullet}_{\mu_\bullet,k}(1-u) \ge \mathrm{q}_k^2(e_\bullet) + \mathrm{q}_k^2(\mu_\bullet) - 2\sqrt{\Sigma_k L_u^2}.$$

For $\mu_\bullet = 0_\bullet$ we have $\mathrm{q}_k^2(\mu_\bullet) = 0$ and $\Sigma_k = \mathrm{q}_k^2(e_\bullet^2)$, hence, we immediately obtain (A.1.1) and (A.1.2). For arbitrary $\mu_\bullet \in \ell^2$ we have $\Sigma_k \leq \mathrm{q}_k^2(e_\bullet^2) + 2\mathrm{q}_k^2(\mu_\bullet)\mathrm{m}_k(e_\bullet^2)$ and, therefore, using $\sqrt{x+y} \leq \sqrt{x} + \sqrt{y}$ and $2\sqrt{xy} \leq cx + c^{-1}x$ for $x, y, c \in \mathbb{R}_+$ with $c = 10$ it follows

$$
\begin{aligned}
2\sqrt{\Sigma_k L_u^2} &\leq 2\sqrt{2\mathrm{q}_k^2(\mu_\bullet)\mathrm{m}_k(e_\bullet^2)L_u^2} + 2\sqrt{\mathrm{q}_k^2(e_\bullet^2)L_u^2} \\
&\leq \frac{1}{5}\mathrm{q}_k^2(\mu_\bullet) + 10\mathrm{m}_k(e_\bullet^2)L_u^2 + 2\sqrt{\mathrm{q}_k^2(e_\bullet^2)L_u^2} \\
&\leq \frac{1}{5}\mathrm{q}_k^2(\mu_\bullet) + (10L_u^2 + 2L_u)\mathrm{q}_k^2(e_\bullet^2),
\end{aligned}
$$

which implies (A.1.3) and completes the proof. □

A.2 Balancing Lemma

The next lemma shows how to balance a monotonically non-increasing sequence with two monotonically non-decreasing sequences, Figure A.2 illustrates the assertion of the lemma. It is needed in many places in this chapter, since we derive upper bounds which are the maximum of two balanced radii with the same decreasing bias term and we want to determine the overall minimising dimension.

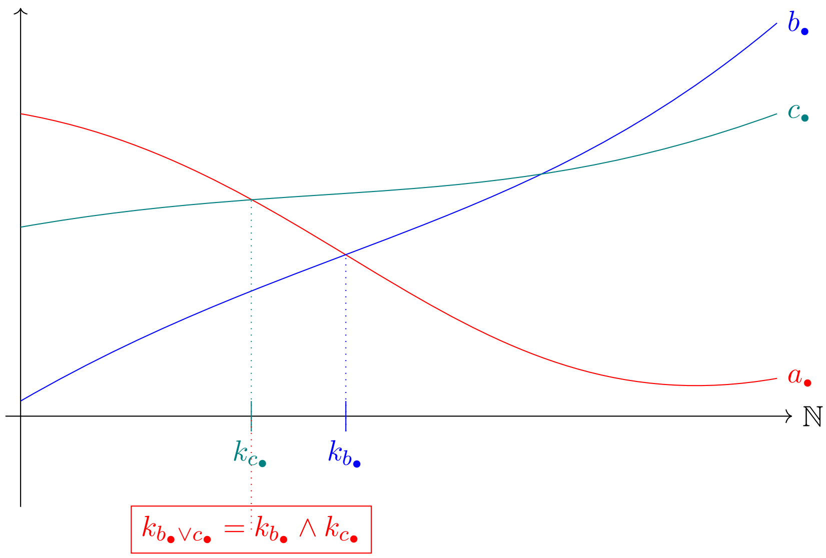


Figure A.1: Illustration of Lemma A.2.1

Lemma A.2.1 (Balancing lemma). Let $a_\bullet \in \mathbb{R}_+^{\mathbb{N}}$ be a monotonically non-increasing sequence, let $b_\bullet, c_\bullet \in \mathbb{R}_+^{\mathbb{N}}$ be two monotonically non-decreasing sequences. Let

$$\rho_{b_\bullet} := \min_{k\in\mathbb{N}} \{a_k \vee b_k\} \text{ and } \rho_{c_\bullet} := \min_{k\in\mathbb{N}} \{a_k \vee c_k\}.$$

Then,

$$\rho_{b_\bullet} \vee \rho_{c_\bullet} = \rho_{b_\bullet \vee c_\bullet} := \min_{k\in\mathbb{N}} \{a_k \vee b_k \vee c_k\}.$$

Moreover,

$$k_{b_\bullet} := \arg\min_{k\in\mathbb{N}} \{a_k \vee b_k\} \text{ and } k_{c_\bullet} := \arg\min_{k\in\mathbb{N}} \{a_k \vee c_k\}$$

satisfy

$$k_{b_\bullet} \wedge k_{c_\bullet} = k_{b_\bullet\vee c_\bullet} := \arg\min_{k\in\mathbb{N}} \{a_k \vee b_k \vee c_k\}.$$

Proof of Lemma A.2.1. We start the proof with the observation that

$$\rho_{b_\bullet} \vee \rho_{c_\bullet} \leq \rho_{b_\bullet\vee c_\bullet}.$$

Since for each $k < k_{b_\bullet\vee c_\bullet}$ we have $a_k \vee b_k \vee c_k = a_k > \rho_{b_\bullet\vee c_\bullet} \geq \rho_{b_\bullet} \vee \rho_{c_\bullet}$, we also obtain

$$k_{b_\bullet} \wedge k_{c_\bullet} \geq k_{b_\bullet\vee c_\bullet}. \tag{A.2.1}$$

Since the minimisers of the sets $\{a_k \vee b_k\}$ and $\{a_k \vee c_k\}$ might not be unique but must be consecutive due to the monotonicity, we define $\overline{k_{b_\bullet\vee c_\bullet}}, \overline{k_{b_\bullet}}, \overline{k_{c_\bullet}} \in \mathbb{N}\cup\{\infty\}$ by

$$\begin{aligned}
[\![k_{b_\bullet}, \overline{k}_{b_\bullet}]\!] &:= \{k \in \mathbb{N} : a_k \vee b_k \leq a_m \vee b_m \forall m \in \mathbb{N}\}\\
[\![k_{c_\bullet}, \overline{k}_{c_\bullet}]\!] &:= \{k \in \mathbb{N} : a_k \vee c_k \leq a_m \vee c_m \forall m \in \mathbb{N}\}\\
[\![k_{b_\bullet\vee c_\bullet}, \overline{k}_{b_\bullet\vee c_\bullet}]\!] &:= \{k \in \mathbb{N} : a_k \vee b_k \vee c_k \leq a_m \vee b_m \vee c_m \forall m \in \mathbb{N}\}
\end{aligned}$$

Now we either have $[\![k_{b_\bullet}, \overline{k}_{b_\bullet}]\!] \subseteq [\![k_{b_\bullet\vee c_\bullet}, \overline{k}_{b_\bullet\vee c_\bullet}]\!]$ or $[\![k_{c_\bullet}, \overline{k}_{c_\bullet}]\!] \subseteq [\![k_{b_\bullet\vee c_\bullet}, \overline{k}_{b_\bullet\vee c_\bullet}]\!]$, because of (A.2.1) and since the non-trivial case $\overline{k}_{b_\bullet\vee c_\bullet} < \infty$ implies for $k := \overline{k}_{b_\bullet\vee c_\bullet} + 1$ that $\rho_{b_\bullet} \vee \rho_{c_\bullet} \leq \rho_{b_\bullet\vee c_\bullet} < b_k \vee c_k = \{a_k \vee b_k\} \vee \{a_k \vee c_k\}$. Without loss of generality let us assume $[\![k_{b_\bullet}, \overline{k}_{b_\bullet}]\!] \subseteq [\![k_{b_\bullet\vee c_\bullet}, \overline{k}_{b_\bullet\vee c_\bullet}]\!]$. Note that there exits a $k \in [\![k_{b_\bullet}, k_{b_\bullet\vee c_\bullet}[\![$ if and only if $\rho_{b_\bullet} < a_k = a_k \vee b_k \leq a_k \vee b_k \vee c_k = \rho_{b_\bullet\vee c_\bullet}$, which in turn implies $\rho_{b_\bullet\vee c_\bullet} = a_k \vee c_k$ for all $k \in [\![k_{b_\bullet}, k_{b_\bullet\vee c_\bullet}[\![$. We distinguish the two cases
(a) $\rho_{b_\bullet} = \rho_{b_\bullet\vee c_\bullet}$
(b) $\rho_{b_\bullet} < \rho_{b_\bullet\vee c_\bullet}$
Firstly, consider (a) which implies $k_{b_\bullet} = k_{b_\bullet\vee c_\bullet}$. Consequently, $\rho_{b_\bullet} \vee \rho_{c_\bullet} \leq \rho_{b_\bullet\vee c_\bullet} = \rho_{b_\bullet} = \rho_{b_\bullet} \vee \rho_{c_\bullet}$ and $k_{b_\bullet} \wedge k_{c_\bullet} \geq k_{b_\bullet} = k_{b_\bullet} \wedge k_{c_\bullet}$, which implies the assertion.
Next, consider (b) which implies $k_{b_\bullet} \geq k_{b_\bullet\vee c_\bullet}$, where $\rho_{b_\bullet\vee c_\bullet} = a_k \vee c_k$ for all $k \in [\![k_{b_\bullet\vee c_\bullet}, k_{b_\bullet}[\![$. Moreover, for all $k \in [\![k_{b_\bullet}, \overline{k}_{b_\bullet}]\!]$ we have $a_k \vee b_k = \rho_{b_\bullet} \leq \rho_{b_\bullet\vee c_\bullet} = a_k \vee b_k \vee c_k$, which in turn implies $\rho_{b_\bullet\vee c_\bullet} = c_k = a_k \vee c_k$ for all $k \in [\![k_{b_\bullet}, \overline{k}_{b_\bullet}]\!]$. Consequently, $a_k \vee c_k = \rho_{b_\bullet\vee c_\bullet}$ for all $k \in [\![k_{b_\bullet\vee c_\bullet}, \overline{k}_{b_\bullet}]\!]$ and $a_{\overline{k}_{b_\bullet}} \vee c_{\overline{k}_{b_\bullet}} \leq c_k = a_k \vee c_k$ for all $k \geq \overline{k}_{b_\bullet}$. Since $\rho_{c_\bullet} \leq \rho_{b_\bullet\vee c_\bullet} < a_k = a_k \vee c_k$ for all $k < k_{b_\bullet\vee c_\bullet}$ it follows $\rho_{c_\bullet} = \rho_{b_\bullet\vee c_\bullet}$ and $k_{c_\bullet} = k_{b_\bullet\vee c_\bullet}$, which in turn implies the claim $\rho_{b_\bullet} \vee \rho_{c_\bullet} = \rho_{b_\bullet\vee c_\bullet}$ and $k_{b_\bullet} \wedge k_{c_\bullet} = k_{b_\bullet\vee c_\bullet}$ and completes the proof. □

A.3 Calculations for the χ^2-divergence

Recall that by $\mathbb{P}^{\varepsilon_\bullet}_{\theta_\bullet}$ we denote the probability distribution of a Gaussian sequence with independent components with mean sequence $\theta_\bullet$ and variance sequence $\varepsilon^2_\bullet$.

Lemma A.3.1 (χ^2-divergence). Let $\mathcal{S}$ be an arbitrary index set with $|\mathcal{S}| = N \in \mathbb{N}$. For $s \in \mathcal{S}$ let $\kappa^s \in \mathbb{N}$, $\theta^s_\bullet \in \ell^2$ and $\nu_\bullet \in \ell^\infty$. For the mixing measure

$$\mathbb{P}_\mu := \frac{1}{N}\sum_{s\in\mathcal{S}} \frac{1}{2^{\kappa^s}} \sum_{\tau\in\{\pm\}^{\kappa^s}} \mathbb{P}^{\varepsilon_\bullet}_{\nu_\bullet\theta^{s,\tau}_\bullet} \qquad \text{with } \theta^{s,\tau}_\bullet = (\tau_j\theta^s_j \mathbb{1}_{\{j\in[\![\kappa^s]\!]\}})_{j\in\mathbb{N}}$$

and $\mathbb{P}_0 := \mathbb{P}^{\varepsilon_\bullet}_{\nu_\bullet \theta^\circ_\bullet}$ the χ^2-divergence satisfies

$$\chi^2(\mathbb{P}_0, \mathbb{P}_\mu) \leq \frac{1}{N^2} \sum_{s,t \in \mathcal{S}} \exp\left(\frac{1}{2} \mathrm{q}^2_{\kappa^s \wedge \kappa^t}(\nu^2_\bullet \tilde{\theta}^s_\bullet \tilde{\theta}^t_\bullet / \varepsilon^2_\bullet)\right) - 1,$$

where $\tilde{\theta}^s_\bullet = \theta^s_\bullet - \theta^\circ_\bullet$.

Proof of Lemma A.3.1. Without loss of generality we assume $\theta^\circ_\bullet = 0_\bullet$ (which is possible since $\nu_\bullet$ is fixed). Inspecting the calculations in the direct Gaussian sequence space model with coordinate-wise constant noise levels by Baraud [2002] (proof of Theorem 1) (compare also the calculations in Lemma B.1.1) it is readily seen that for any $z_\bullet = (z_j)_{j \in \mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ the likelihood ratio is given by

$$\frac{\mathrm{d}\mathbb{P}_\mu}{\mathrm{d}\mathbb{P}_0}(z_\bullet) = \frac{1}{N} \sum_{s \in \mathcal{S}} \exp\left(-\frac{1}{2} \mathrm{q}^2_{\kappa^s}(\nu_\bullet \tilde{\theta}^s_\bullet / \varepsilon_\bullet)\right) \prod_{j=1}^{\kappa^s} \frac{1}{2} \left(\exp\left(-\frac{\nu_j \widetilde{\theta}^s_j z_j}{\varepsilon^2_j}\right) + \exp\left(\frac{\nu_j \widetilde{\theta}^s_j z_j}{\varepsilon^2_j}\right)\right).$$

By taking the expectation of the squared likelihood ratio with respect to $\mathbb{P}_0$ we obtain

$$\begin{aligned}\mathbb{E}_0 \left(\frac{\mathrm{d}\mathbb{P}_\mu}{\mathrm{d}\mathbb{P}_0}(Z_\bullet)\right)^2 &= \frac{1}{N^2} \sum_{s,t \in \mathcal{S}} \prod_{j=1}^{\kappa^s \wedge \kappa^t} \frac{1}{2} \left(\exp\left(-\frac{\nu^2_j \widetilde{\theta}^s_j \widetilde{\theta}^t_j}{\varepsilon_j}\right) + \exp\left(\frac{\nu^2_j \widetilde{\theta}^s_j \widetilde{\theta}^t_j}{\varepsilon_j}\right)\right) \\ &= \frac{1}{N^2} \sum_{s,t \in \mathcal{S}} \prod_{j=1}^{\kappa^s \wedge \kappa^t} \cosh\left(\frac{\nu^2_j \widetilde{\theta}^s_j \widetilde{\theta}^t_j}{\varepsilon^2_j}\right),\end{aligned}$$

where $Z_\bullet$ is a random variable with distribution $\mathbb{P}_0$. Exploiting the elementary inequality $\cosh(x) \leq \exp(x^2/2)$, $x \in \mathbb{R}$ and the definition of the χ^2-divergence completes the proof. □

A.4 Mixtures

The next example provides a heuristic explanation in a simplified setting why taking mixtures helps to obtain better results when proving lower bounds.

Example A.4.1 (Mixtures help). We consider a parametric testing problem. Let $\mathbb{P}_\theta$ be the probability measure of a normal distribution $\mathcal{N}(\theta, 1)$ and let $\theta \in \mathbb{R}$ be the quantity of interest. When analysing the complexity of the testing problem

$$H_0 : \theta = 0 \qquad \text{against} \qquad H_1 : |\theta| \geq \rho,$$

we compare the distance between the sets of probability distributions $\{\mathbb{P}_0\}$ and $\{\mathbb{P}_\theta : |\theta| \geq \rho\}$. Standard techniques for proving lower bounds in this setting involve the total variation distance between $\mathbb{P}_0$ and a mixture $\mathbb{P}_\mu$ with $\mathbb{P}_\mu(A) := \int \mathbb{P}_\theta(A)\mathrm{d}\mu(\theta)$ for measurable sets A and a mixing measure μ supported on the parameter set $\{\theta : |\theta| \geq \rho\}$ of the alternative. This example shows that indeed taking a mixing measure over several parameters of the alternative instead of taking just one element of the alternative gives a better lower bound (i.e. a smaller total variation distance). In fact, we have

$$\mathrm{TV}(\mathbb{P}_0, \mathbb{P}_\theta) = \sqrt{\tfrac{1}{2\pi}}\, |\theta| + \mathcal{O}(\theta^2) \qquad \text{as } \theta \longrightarrow 0.$$

Choosing $\mu := \frac{1}{2}(\delta_\theta + \delta_{-\theta})$, i.e. $\mathbb{P}_\mu = \frac{1}{2}(\mathcal{N}(\theta, 1) + \mathcal{N}(-\theta, 1))$ yields

$$\mathrm{TV}(\mathbb{P}_0, \mathbb{P}_\mu) = C\theta^2 + \mathcal{O}(\theta^4) \qquad \text{as } \theta \longrightarrow 0.$$

for some positive constant $C > 0$.

Proof. We first recall Taylor's Theorem with a remainder in Lagrange form. For a $(k+1)$-times differentiable function f we have for $x \in \mathbb{R}$, $\theta \in \mathbb{R}^+$ and some $\xi \in [x, x+\theta]$

$$f(x+\theta) = f(x) + \theta f^{(1)}(x) + \frac{\theta^2}{2!} f^{(2)}(x) + \cdots + \frac{\theta^k}{k!} f^{(k)}(x) + \frac{\theta^{k+1}}{(k+1)!} f^{(k+1)}(\xi).$$

This implies

$$\begin{aligned} f(x+\theta) - f(x) &= \theta f^{(1)}(x) + \frac{\theta^2}{2} f^{(2)}(\xi_0) \\ f(x+\theta) + f(x-\theta) - 2f(x) &= \theta^2 f^{(2)}(x) + \frac{\theta^4}{12}\left(f^{(4)}(\xi_1) + f^{(4)}(\xi_2)\right) \end{aligned}$$

for some $\xi_0, \xi_1 \in [x, x+\theta]$, $\xi_2 \in [x-\theta, x]$. The second equation is called the **second symmetric derivative**. Moreover, if f is the density of the standard normal distribution, we have

$$\begin{aligned} f(x) &= \tfrac{1}{\sqrt{2\pi}} \exp\left(-\tfrac{x^2}{2}\right) \\ f^{(1)}(x) &= -x f(x) \\ f^{(2)}(x) &= (x^2 - 1) f(x) \\ f^{(3)}(x) &= (2x - x^3) f(x). \end{aligned}$$

Then, for some $\xi_x \in [x, x+\theta]$, we obtain

$$\begin{aligned} \mathrm{TV}(\mathbb{P}_0, \mathbb{P}_\theta) &= \frac{1}{2}\int |f(x+\theta) - f(x)|\,\mathrm{d}x = \frac{1}{2}\int \left|\theta f^{(1)}(x) + \frac{\theta^2}{2} f^{(2)}(\xi_x)\right| \mathrm{d}x \\ &= \frac{1}{2}\int \left|\theta f^{(1)}(x)\right| \mathrm{d}x + \mathcal{O}(\theta^2) \\ &= \frac{1}{2}|\theta| \int |x|\, f(x)\mathrm{d}x + \mathcal{O}(\theta^2) = \frac{1}{2}|\theta|\sqrt{\frac{2}{\pi}} + \mathcal{O}(\theta^2) \end{aligned}$$

since $\int |x|\, f(x)\mathrm{d}x = 2\int_{[0,\infty)} x f(x)\mathrm{d}x = \frac{2}{\sqrt{2\pi}} e^{-x^2/2} \,|_{x=0}^{x=\infty} = \sqrt{\frac{2}{\pi}}$, which proves the first assertion. For the second assertion we use the representation for the second symmetric derivative derived above and obtain for $\xi_1 = \xi_{1,x} \in [x, x+\theta]$ and $\xi_2 = \xi_{2,x} \in [x-\theta, x]$

$$\begin{aligned} \mathrm{TV}(\tfrac{1}{2}(\mathbb{P}_\theta + \mathbb{P}_{-\theta}), \mathbb{P}_0) &= \frac{1}{2}\int \left|\frac{1}{2}(f(x+\theta) + f(x-\theta)) - f(x)\right| \mathrm{d}x \\ &= \frac{1}{4}\int |f(x+\theta) + f(x-\theta) - f(x)|\,\mathrm{d}x \\ &= \frac{1}{4}\int \left|\theta^2 f^{(2)}(x) + \frac{\theta^4}{12}\left(f^{(4)}(\xi_1) + f^{(4)}(\xi_2)\right)\right| \mathrm{d}x \\ &= \frac{1}{4}\int \left|\theta^2 f^{(2)}(x)\right| \mathrm{d}x + \mathcal{O}(\theta^4) \\ &= \frac{1}{4}\theta^2 \int \left|(x^2 - 1)\right| f(x)\mathrm{d}x + \mathcal{O}(\theta^4) \\ &= C\theta^2 + \mathcal{O}(\theta^4) \end{aligned}$$

with $C := \frac{1}{4}\int \left|(x^2 - 1)\right| f(x)\mathrm{d}x \in (0, \infty)$, which completes the proof. □

Chapter 2

Testing of linear functionals

In this chapter we derive matching upper and lower bounds for the minimax separation radius in a linear functional testing problem for the inverse Gaussian sequence space model. Moreover, we compare linear functional testing to goodness-of-fit testing.

2.1 Linear functional testing

We consider an inverse Gaussian sequence space model, i.e. our observations are given by

$$Y_j = \lambda_j \theta_j + \varepsilon \xi_j, \qquad j \in \mathbb{N},$$

where $\varepsilon > 0$ is the noise level, $\xi_j \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$ is i.i.d. white noise, $\lambda_\bullet = (\lambda_j)_{j\in\mathbb{N}} \in \ell^\infty$ is a known bounded sequence and $\theta_\bullet = (\theta_j)_{j\in\mathbb{N}} \in \ell^2$ is an unknown square summable sequence of interest. For an introduction of the model we refer to (IGSSM). We denote by $\mathbb{P}_{\theta_\bullet}$ the joint law of $(Y_j)_{j\in\mathbb{N}}$ with mean $(\lambda_\bullet\theta_\bullet)$. In this chapter instead of making inference on the signal $\theta_\bullet = (\theta_j)_{j\in\mathbb{N}}$ itself, we aim to make inference on the value of a linear functional $L(\theta_\bullet)$. Let

$$L : \mathcal{D} \longrightarrow \mathbb{R}$$

be a linear functional and $\theta_\bullet^\circ \in \mathcal{D} \subseteq \ell^2$ a benchmark sequence with $L^\circ := L(\theta_\bullet^\circ)$, where $\mathcal{D} = \mathcal{D}(L) \subseteq \ell^2$ denotes the natural domain of L. For a separation radius $\rho > 0$ and a nonparametric class $\Theta \subseteq \mathcal{D}$ we consider the testing problem

$$H_0 : L(\theta_\bullet) = L^\circ \qquad \text{against} \qquad H_1^\rho : |L(\theta_\bullet) - L^\circ| \geq \rho, \theta_\bullet - \theta_\bullet^\circ \in \Theta. \tag{2.1.1}$$

We present the testing task in the form (2.1.1), which is typical for non-parametric minimax testing. Note, however, that in practice when testing $L(\theta_\bullet) = L^\circ$ for a given L° the benchmark $\theta_\bullet^\circ$ is generally not given (and not uniquely identified by $L(\theta_\bullet^\circ) = L^\circ$). Hence, in our proofs we only require that there exists a $\theta_\bullet^\circ \in \Theta$ with $L(\theta_\bullet^\circ) = L^\circ$ and control the error probabilities of a test for any element $\theta_\bullet$ such that there exists such a $\theta_\bullet^\circ$ with $\theta_\bullet^\circ - \theta_\bullet \in \Theta$.
For a test Δ, i.e. a measurable function $\Delta : \mathbb{R}^\mathbb{N} \longrightarrow \{0,1\}$, we define the **maximal risk** corresponding to the testing task (2.1.1) by setting

$$\mathcal{R}\left(\Delta \mid \Theta, \rho\right) := \mathbb{P}_{\theta_\bullet^\circ}(\Delta = 1) + \sup_{\substack{\theta_\bullet - \theta_\bullet^\circ \in \Theta \\ |L(\theta_\bullet) - L^\circ| \geq \rho}} \mathbb{P}_{\theta_\bullet}(\Delta = 0).$$

We measure the difficulty of the testing task by taking the smallest risk obtainable by any test, that is, we define the **minimax risk** $\mathcal{R}\left(\Theta, \rho\right) := \inf_\Delta \mathcal{R}\left(\Delta \mid \Theta, \rho\right)$. We call $\rho = \rho(\Theta)$ a minimax radius of testing if for all $\alpha \in (0,1)$ there exist constants $\underline{A}_\alpha, \overline{A}_\alpha \in \mathbb{R}_+$ such that

(i) for all $A \geq \overline{A}_\alpha$ we have $\mathcal{R}(\Theta, A\rho) \leq \alpha$, (upper bound)

(ii) for all $A \leq \underline{A}_\alpha$ we have $\mathcal{R}(\Theta, A\rho) \geq 1 - \alpha$. (lower bound)

The goal of this chapter is to determine a radius of testing over an ℓ^2-ellipsoid Θ for an arbitrary linear functional L. Let us formalize the kind of alternatives we consider. For a monotonically non-increasing sequence $a_\bullet$ and a positive radius $\mathrm{R} > 0$ we define the non-parametric class of sequences

$$\Theta = \Theta_{a_\bullet}^{\mathrm{R}} := \left\{ \vartheta_\bullet \in \ell^2 : \sum_{j \in \mathbb{N}} \vartheta_j^2 a_j^{-2} \leq \mathrm{R}^2 \right\}$$

and assume that $\theta_\bullet^\circ \in \Theta_{a_\bullet}^{\mathrm{R}}$.

Linear functionals. By the Riesz representation theorem (see Theorem V.3.6. in Werner [2006]) there exists a sequence $L_\bullet = (L_j)_{j \in \mathbb{N}} \in \ell^2$ such that any continuous (i.e. bounded) linear functional $L : \ell^2 \longrightarrow \mathbb{R}$ can be represented as $L(\theta_\bullet) = \langle L_\bullet, \theta_\bullet \rangle_{\ell^2} = \sum_{j \in \mathbb{N}} L_j \theta_j$ for all $\theta_\bullet \in \ell^2$. We point out that we do not need the square summability of the coefficients $L_\bullet \in \ell^2$ of the linear functional for our testing method, since we base our test statistic on a finite number of coefficients anyway. Due to the regularity assumption, it is sufficient if $\sum_{j \in \mathbb{N}} L_j^2 a_j^2 < \infty$, since then

$$\begin{aligned}
\sum_{j \in \mathbb{N}} |L_j \theta_j| &\leq \sum_{j \in \mathbb{N}} \left| L_j (\theta_j - \theta_j^\circ) \right| + \sum_{j \in \mathbb{N}} \left| L_j \theta_j^\circ \right| \\
&\leq \left(\sum_{j \in \mathbb{N}} L_j^2 a_j^2 \right)^{1/2} \left(\sum_{j \in \mathbb{N}} a_j^{-2} (\theta_j - \theta_j^\circ)^2 \right)^{1/2} + \left(\sum_{j \in \mathbb{N}} L_j^2 a_j^2 \right)^{1/2} \left(\sum_{j \in \mathbb{N}} a_j^{-2} (\theta_j^\circ)^2 \right)^{1/2} \\
&\leq 2\mathrm{R} \left(\sum_{j \in \mathbb{N}} L_j^2 a_j^2 \right)^{1/2} < \infty.
\end{aligned}$$

Hence, we consider any linear functional which has a representation of the form

$$L(\theta_\bullet) = \langle L_\bullet, \theta_\bullet \rangle_{\ell^2} = \sum_{j \in \mathbb{N}} L_j \theta_j \qquad \text{with} \qquad \sum_{j \in \mathbb{N}} L_j^2 a_j^2 < \infty. \tag{2.1.2}$$

Thus, we are able to cover a larger class of linear functionals. Let us give some examples of such functionals. Let $f \in \mathscr{L}^2[0,1)$ be a square integrable real-valued function, denote by $f_\bullet := (f_j)_{j \in \mathbb{N}}$ its coefficients in some basis $(b_j)_{j \in \mathbb{N}}$ of $\mathscr{L}^2[0,1)$. Linear functionals acting on (subsets of) $\mathscr{L}^2[0,1)$, such as point evaluation, average values or weighted averages can be represented as linear functionals acting on the coefficients in ℓ^2. By $\mathbb{D}$ and $\mathcal{D}$ we denote the respective (natural) domains of the functionals.

- **Point evaluation**. Let $t_\circ \in [0,1)$. Define $\mathbb{L} : \mathscr{L}^2[0,1) \supseteq \mathbb{D} \longrightarrow \mathbb{R}$ by

  $$\mathbb{L} f = f(t_\circ).$$

  We have $f(t_\circ) = \sum_{j \in \mathbb{N}} f_j b_j(t_\circ) =: \sum_{j \in \mathbb{N}} L_j f_j =: L(f_\bullet)$ with $L_j := b_j(t_\circ)$, i.e. the corresponding linear functional $L : \ell^2 \supseteq \mathcal{D} \longrightarrow \mathbb{R}$ satisfies $\mathbb{L} f = L(f_\bullet)$ with domain $\mathcal{D} = \left\{ f_\bullet : \sum_{j \in \mathbb{N}} |f_j| \, |b_j(t_\circ)| < \infty \right\}$. In particular, this linear functional is unbounded.

- **Averages**. Let $c \in [0,1)$. The average (up to c) functional $\mathbb{L} : \mathscr{L}^2[0,1) \longrightarrow \mathbb{R}$ is given by

  $$\mathbb{L} f = \int_0^c f(t) \mathrm{d}t.$$

  We have $\int_0^c f(t)\mathrm{d}t = \sum_{j \in \mathbb{N}} f_j \int_0^c b_j(t)\mathrm{d}t =: \sum_{j \in \mathbb{N}} L_j f_j =: L(f_\bullet)$ with $L_j := \int_0^c b_j(t)\mathrm{d}t$, i.e. the corresponding linear functional $L : \ell^2 \longrightarrow \mathbb{R}$ satisfies $\mathbb{L} f = L(f_\bullet)$.

▶ **Weighted Averages**. Let $\omega \in \mathscr{L}^2[0,1)$. The weighted average functional $\mathbb{L} : \mathscr{L}^2[0,1) \longrightarrow \mathbb{R}$ is given by

$$\mathbb{L}f = \int_0^1 \omega(t)f(t)\mathrm{d}t.$$

We have $\int_0^1 \omega(t)f(t)\mathrm{d}t = \sum_{j\in\mathbb{N}} f_j \int_0^1 \omega(t)b_j(t)\mathrm{d}t =: \sum_{j\in\mathbb{N}} L_j f_j =: L(f_\bullet)$ with coefficients $L_j := \int_0^1 \omega(t)b_j(t)\mathrm{d}t$, i.e. the corresponding functional $L : \ell^2 \longrightarrow \mathbb{R}$ satisfies $\mathbb{L}f = L(f_\bullet)$.

Related literature. Statistical estimation of linear functional dates back to the 80s. Early work includes Ibragimov and Khas' minskii [1985] and Ibragimov and Khas' minskii [1988] (in Gaussian noise models) and Ibragimov and Khas' minskii [1989] (in a density observation model). Goldenshluger and Pereverzev [2000] consider the adaptive estimation of linear functionals from indirect white noise observations. The results are extended to Hilbert scales in Goldenshluger and Pereverzev [2003]. Let us mention some further work in sequence space models. Adaptive estimation over Besov balls under $\mathscr{L}^p$-loss is treated e.g. in Laurent et al. [2008]. The series of papers Cai and Low [2003], Cai and Low [2004], Cai and Low [2005b], Cai and Low [2005c] covers the adaptive estimation of linear functionals over convex and non-convex function classes characterised in terms of a modulus of continuity. Under sparsity assumptions estimation of linear functionals is considered more recently in Collier et al. [2018], treating both adaptation to the smoothness index and the noise level in a direct Gaussian sequence space model. For a specific linear functional only (i.e. $L(\theta_\bullet) := \sum_{j\in\mathbb{N}} \theta_j$) Golubev [2020] considers adaptive estimation, comparing adaptive choices of a cut-off parameter. Butucea and Comte [2009] consider linear functional estimation in a convolution model.
Interestingly, also direct approaches to estimation of linear functionals are investigated, for instance in Mathé and Pereverzev [2002]. Instead of aiming to solve the observation equation $Y = Th + \varepsilon\xi$ for h and then applying the linear functional $Lh = \langle l, h\rangle$ (called the *solution-functional approach*), they search for a (regularized) solution f of $T^*f = l$, where T^* is the adjoint operator of T, then $Lh = \langle l, h\rangle = \langle T^*f, h\rangle = \langle f, Th\rangle$ and therefore their estimation technique uses the observations of Th directly (called the *data-functional approach*). Although we do not use this technique in this chapter it is worth mentioning since it is a similar to the direct and indirect approaches for testing considered in this thesis.

Outline of this chapter. We derive an upper bound in Section 2.2 and a matching lower bound in Section 2.3. We illustrate the resulting minimax separation radii for typical ill-posedness and smoothness cases in Illustration 2.4.1. Moreover, we compare linear functional testing to goodness-of-fit testing in Section 2.4 and discuss adaptation (Remark 2.4.3).

2.2 Upper bound

Definition of the test. Our test is based on a truncated plug-in estimator of the distance $|L(\theta_\bullet) - L^\circ|$. Due to the representation (2.1.2) for a cut-off dimension $k \in \mathbb{N}$ we suggest the test statistic

$$\hat{l}_k := \sum_{j\in[\![k]\!]} L_j \frac{Y_j}{\lambda_j} - L^\circ$$

as a truncated estimator of $L(\theta_\bullet) - L^\circ$. For $\alpha \in (0,1)$ we define the threshold

$$\tau_k(\alpha) := \mathrm{q}_{\alpha/2}\varepsilon\sqrt{\sum_{j\in[\![k]\!]} \frac{L_j^2}{\lambda_j^2}} + \sqrt{\mathrm{R}^2 \sum_{j>k} L_j^2 a_j^2} \qquad (2.2.1)$$

where $\mathfrak{q}_\alpha$ denotes the $(1-\alpha)$-quantile of a standard normal distribution, i.e. $\mathbb{P}(Z \le \mathfrak{q}_\alpha) = 1-\alpha$ for $Z \sim \mathcal{N}(0,1)$. Finally, we introduce the test

$$\Delta_{k,\alpha} := \mathbb{1}_{\{|\hat{l}_k| > \tau_k(\alpha)\}}. \tag{2.2.2}$$

Proposition 2.2.1 (Quantiles of the test statistic). Let $\alpha, \beta \in (0,1)$. Denote by $\mathfrak{q}_\alpha$ the $(1-\alpha)$-quantile of a standard normal distribution. Let $k \in \mathbb{N}$ and consider the threshold $\tau_k(\alpha)$ defined in (2.2.1) and a benchmark function $\theta^\circ_\bullet \in \Theta^{\mathrm{R}}_{a_\bullet}$ with $L(\theta^\circ_\bullet) = L^\circ$.

(i) **(α-level)** We have

$$\mathbb{P}_{\theta^\circ_\bullet}(\Delta_{k,\alpha} = 1) = \mathbb{P}_{\theta^\circ_\bullet}\left(\left|\hat{l}_k\right| > \tau_k(\alpha)\right) \le \alpha.$$

(ii) **($(1-\beta)$-powerful)** Let $\theta_\bullet$ satisfy $\theta_\bullet - \theta^\circ_\bullet \in \Theta^{\mathrm{R}}_{a_\bullet}$ and

$$|L(\theta_\bullet) - L^\circ| \ge (\mathfrak{q}_{\alpha/2} - \mathfrak{q}_{1-\beta})\varepsilon\sqrt{\sum_{j\in[\![k]\!]}\frac{L_j^2}{\lambda_j^2}} + 3\sqrt{\mathrm{R}^2\sum_{j>k} L_j^2 a_j^2} =: \rho$$

then we have

$$\mathbb{P}_{\theta_\bullet}(\Delta_{k,\alpha} = 0) = \mathbb{P}_{\theta_\bullet}\left(\left|\hat{l}_k\right| \le \tau_k(\alpha)\right) \le \beta.$$

Proof of Proposition 2.2.1. Note that for $Y_j \sim \mathcal{N}(\lambda_j\theta_j, \varepsilon^2)$ our test statistic $\hat{l}_k$ follows a normal distribution with mean $\mu_{\theta_\bullet,k} := \sum_{j\in[\![k]\!]} L_j(\theta_j - \theta_j^\circ) - \sum_{j>k} L_j\theta_j^\circ$ and variance $\sigma_k^2 := \varepsilon^2\sum_{j\in[\![k]\!]}\frac{L_j^2}{\lambda_j^2}$.

(i) Hence, we can rewrite the probability

$$\begin{aligned}\mathbb{P}_{\theta^\circ_\bullet}\left(\left|\hat{l}_k\right| > \tau_k(\alpha)\right) &= \mathbb{P}_{\theta^\circ_\bullet}\left(\hat{l}_k > \tau_k(\alpha)\right) + \mathbb{P}_{\theta^\circ_\bullet}\left(\hat{l}_k < -\tau_k(\alpha)\right)\\ &\le 2\mathbb{P}_{\theta^\circ_\bullet}\left(Z > \frac{\tau_k(\alpha) - |\mu_{\theta^\circ_\bullet,k}|}{\sigma_k}\right),\end{aligned}$$

where $Z \sim \mathcal{N}(0,1)$. Note that for $\theta_\bullet = \theta^\circ_\bullet$ we have $\mu_{\theta^\circ_\bullet,k} = -\sum_{j>k} L_j\theta_j^\circ$. We point out that $\theta^\circ_\bullet$ is not given in practice (only L° is!) and it is not necessarily uniquely determined by $L(\theta^\circ_\bullet) = L^\circ$. Hence, in general the bias term $\mu_{\theta^\circ_\bullet,k}$ is unknown but easily bounded using the Cauchy-Schwarz inequality

$$|\mu_{\theta^\circ_\bullet,k}| \le \sum_{j>k}\left|L_j\theta_j^\circ\right| \le \left(\sum_{j>k}(\theta_j^\circ)^2 a_j^{-2}\right)^{1/2}\left(\sum_{j>k} L_j^2 a_j^2\right)^{1/2} \le \mathrm{R}\left(\sum_{j>k} L_j^2 a_j^2\right)^{1/2} := \mathbb{b}_k$$

due to the regularity assumption $\theta^\circ_\bullet \in \Theta^{\mathrm{R}}_{a_\bullet}$. Therefore, inserting the definition of the threshold yields the desired upper bound.

$$\mathbb{P}_{\theta^\circ_\bullet}\left(\left|\hat{l}_k\right| > \tau_k(\alpha)\right) \le 2\mathbb{P}_{\theta^\circ_\bullet}\left(Z > \frac{\tau_k(\alpha) - \mathbb{b}_k}{\sigma_k}\right) = 2\mathbb{P}_{\theta^\circ_\bullet}\left(Z > \mathfrak{q}_{\alpha/2}\right) = 2\frac{\alpha}{2} = \alpha.$$

(ii) Let $\theta_\bullet - \theta^\circ_\bullet \in \Theta^{\mathrm{R}}_{a_\bullet}$, due to the Cauchy-Schwarz inequality we have

$$\left|\sum_{j>k} L_j(\theta_j - \theta_j^\circ)\right| \le \left(\sum_{j>k} L_j^2 a_j^2\right)^{1/2}\left(\sum_{j>k}\left(\theta_j^\circ - \theta_j\right)^2 a_j^{-2}\right)^{1/2} \le \mathrm{R}\left(\sum_{j>k} L_j^2 a_j^2\right)^{1/2}$$

and in the same manner

$$\left|\sum_{j>k} L_j\theta^\circ_j\right| \le \mathrm{R}\left(\sum_{j>k} L_j^2 a_j^2\right)^{1/2}.$$

First consider the case $L(\theta_\bullet) - L^\circ \ge \rho$. Hence, $|L(\theta_\bullet) - L^\circ| \ge \rho$ implies

$$\begin{aligned}\mu_{\theta_\bullet,k} &\ge \rho - \sum_{j>k} L_j\theta^\circ_j - \sum_{j>k} L_j(\theta_j - \theta^\circ_j)\\ &\ge (\mathrm{q}_{\alpha/2} - \mathrm{q}_{1-\beta})\varepsilon\sqrt{\sum_{j\in[\![k]\!]} \frac{L_j^2}{\lambda_j^2}} + \sqrt{\mathrm{R}^2\sum_{j>k} L_j^2 a_j^2}.\end{aligned}$$

Rearranging the last inequality and recalling the definition of the threshold $\tau_k(\alpha)$ in (2.2.1) yields

$$\frac{\tau_k(\alpha) - \mu_{\theta_\bullet,k}}{\sigma_k} \le \mathrm{q}_{1-\beta}.$$

Therefore, we can bound the probability

$$\begin{aligned}\mathbb{P}_{\theta_\bullet}\left(\left|\hat{l}_k\right| \le \tau_k(\alpha)\right) &= \mathbb{P}_{\theta_\bullet}\left(\hat{l}_k \in [-\tau_k(\alpha), \tau_k(\alpha)]\right)\\ &= \mathbb{P}_{\theta_\bullet}\left(\frac{\hat{l}_k - \mu_{\theta_\bullet,k}}{\sigma_k} \in \left[-\frac{\tau_k(\alpha) - \mu_{\theta_\bullet,k}}{\sigma_k}, \frac{\tau_k(\alpha) - \mu_{\theta_\bullet,k}}{\sigma_k}\right]\right)\\ &\le \mathbb{P}_{\theta_\bullet}\left(Z \le \frac{\tau_k(\alpha) - \mu_{\theta_\bullet,k}}{\sigma_k}\right) \le \mathbb{P}_{\theta_\bullet}\left(Z \le \mathrm{q}_{1-\beta}\right) = \beta\end{aligned}$$

for a standard normal random variable $Z \sim \mathcal{N}(0,1)$, which proves the assertion. The other case $L(\theta_\bullet) - L^\circ \le -\rho$ follows analogously by considering the bound

$$\mathbb{P}_{\theta_\bullet}\left(\hat{l}_k \in [-\tau_k(\alpha), \tau_k(\alpha)]\right) \le \mathbb{P}_{\theta_\bullet}\left(\hat{l}_k \ge -\tau_k(\alpha)\right)$$

and proceeding as above.

□

For a truncation dimension $k \in \mathbb{N}$ let us define the separation radius

$$\rho_k := \sqrt{\varepsilon^2\sum_{j\in[\![k]\!]} \frac{L_j^2}{\lambda_j^2} + \sum_{j>k} L_j^2 a_j^2}.$$

The part $\sqrt{\sum_{j>k} L_j^2 a_j^2}$ is a typical bias term, which decreases if the dimension parameter k increases. On the other hand the variance part $\varepsilon\sqrt{\sum_{j\in[\![k]\!]} \frac{L_j^2}{\lambda_j^2}}$ increases with k. Proposition 2.2.1 in particular shows that $\Delta_{k,\alpha}$ is a level-α-test that is $(1-\beta)$-powerful over $2(\mathrm{q}_{\alpha/2}-\mathrm{q}_{1-\beta}+3\mathrm{R})\rho_k$-separated alternatives (since $\sqrt{x}+\sqrt{y} \le 2\sqrt{x+y}$ for all $x, y \ge 0$). The radius ρ_k can be optimised with respect to the cut-off dimension k. In the case of a monotonically decreasing sequence $\lambda_\bullet = (\lambda_j)_{j\in\mathbb{N}}$, which we assume from hereon, the optimal dimension is given by the simplified expression

$$\kappa_\star := \sup\left\{k \in \mathbb{N} : \frac{\varepsilon^2}{\lambda_k^2} < a_k^2\right\}$$

and the corresponding minimal radius is

$$\rho_\star := \min_{k\in\mathbb{N}}\left\{\sqrt{\varepsilon^2\sum_{j\in[\![k]\!]} \frac{L_j^2}{\lambda_j^2} + \sum_{j>k} L_j^2 a_j^2}\right\}. \tag{2.2.3}$$

The next proposition establishes $\rho_\star$ as an upper bound based on Proposition 2.2.1

Proposition 2.2.2 (Upper bound for the radius of testing). For $\alpha \in (0,1)$ define

$$\overline{A}_\alpha := 2\left(3\mathrm{R} + \mathrm{q}_{\alpha/4} - \mathrm{q}_{1-\alpha/2}\right).$$

where q_α denotes the $(1-\alpha)$-quantile of a standard normal random variable. Then, for all $A \geq \overline{A}_\alpha$ we obtain

$$\mathcal{R}\left(\Theta^{\mathrm{R}}_{a_\bullet}, A\rho_\star\right) \leq \mathcal{R}\left(\Delta_{\kappa_\star,\alpha/2} \mid \Theta^{\mathrm{R}}_{a_\bullet}, A\rho_\star\right) \leq \alpha,$$

i.e. $\rho_\star$ is an upper bound for the minimax radius of testing.

Proof of Proposition 2.2.2. The assertion follows from Proposition 2.2.1 applied to $\beta = \alpha/2$ and the definition of the testing risk. □

2.3 Lower bound

In this section we provide a matching lower bound to the upper bound derived in Proposition 2.2.2. Thus, we establish $\rho_\star$ defined in (2.2.3) as the minimax radius of testing.

Proposition 2.3.1 (Lower bound). For $\alpha \in (0,1)$ we define $\underline{A}_\alpha := \mathrm{R} \wedge \sqrt{\log(1+2\alpha^2)}$. Let $\rho_\star$ be defined in (2.2.3). Then for all $A \leq \underline{A}_\alpha$ we have

$$\mathcal{R}\left(\Theta^{\mathrm{R}}_{a_\bullet}, \underline{A}_\alpha\rho_\star\right) \geq 1-\alpha,$$

i.e. $\rho_\star$ is a lower bound for the minimax radius of testing.

Proof of Proposition 2.3.1. **Reduction step.** Standard reduction techniques show that for any test Δ the testing risk is lower bounded by

$$\mathcal{R}\left(\Delta \mid \Theta^{\mathrm{R}}_{a_\bullet}, \underline{A}_\alpha\rho_\star\right) \geq 1 - \sqrt{\frac{\chi^2(\mathbb{P}_{\theta^\circ_\bullet}, \mathbb{P}_{\theta_\bullet})}{2}}$$

for some sequence $\theta_\bullet$ (called hypothesis) contained in the $\underline{A}_\alpha\rho_\star$-separated alternative, i.e. satisfying $\theta_\bullet - \theta^\circ_\bullet \in \Theta^{\mathrm{R}}_{a_\bullet}$ and $|L(\theta_\bullet) - L^\circ| \geq \underline{A}_\alpha\rho_\star$. Moreover, straight-forward calculations (detailed in Lemma B.1.1) show that

$$\chi^2(\mathbb{P}_{\theta^\circ_\bullet}, \mathbb{P}_{\theta_\bullet}) = \exp\left(\frac{\sum_{j\in\mathbb{N}} \lambda_j^2 \left(\theta^\circ_j - \theta_j\right)^2}{\varepsilon^2}\right) - 1.$$

Construction of the hypothesis $\theta_\bullet$. We define the sequence $\theta_\bullet$ by setting

$$\theta_j := \begin{cases} \xi\varepsilon^2 \frac{L_j}{\lambda_j^2\rho_\star} + \theta^\circ_j & \text{if } j \leq \kappa_\star, \\ \xi a_j^2 \frac{L_j}{\rho_\star} + \theta^\circ_j & \text{if } j > \kappa_\star, \end{cases}$$

with $\xi := \mathrm{R} \wedge \sqrt{\log(1+2\alpha^2)}$, which has the following three properties.

(1) $|L(\theta_\bullet) - L^\circ| \geq c\rho_\star$ (separation)
By construction it follows

$$|L(\theta_\bullet) - L^\circ| = \xi \sum_{j>\kappa_\star} a_j^2 \frac{L_j^2}{\rho_\star} + \xi\varepsilon^2 \sum_{j\leq\kappa_\star} \frac{L_j^2}{\lambda_j^2\rho_\star} = \xi\frac{\rho_\star^2}{\rho_\star} = \xi\rho_\star.$$

(2) $\theta_\bullet - \theta^\circ_\bullet \in \Theta^{\mathrm{R}}_{a_\bullet}$ (smoothness)
Note that $\frac{\varepsilon^2}{\lambda_j^2} a_j^{-2} \leq 1$ for all $j \leq \kappa_\star$. Hence,

$$\sum_{j\in\mathbb{N}} a_j^{-2} \left(\theta_j - \theta_j^\circ\right)^2 = \xi^2 \sum_{j\leq\kappa_\star} \varepsilon^4 \frac{L_j^2}{\lambda_j^4 \rho_\star^2} a_j^{-2} + \xi^2 \sum_{j>\kappa_\star} a_j^4 \frac{L_j^2}{\rho_\star^2} a_j^{-2}$$
$$\leq \xi^2 \sum_{j\leq\kappa_\star} \varepsilon^2 \frac{L_j^2}{\lambda_j^2 \rho_\star^2} + \xi^2 \sum_{j>\kappa_\star} a_j^2 \frac{L_j^2}{\rho_\star^2}$$
$$\leq \xi^2 \frac{\rho_\star^2}{\rho_\star^2} = \xi^2 \leq \mathrm{R}^2,$$

since for $x, y > 0$ we have $x^2 + y^2 \leq (x+y)^2$.

(3) $\frac{\sum_{j\in\mathbb{N}} \lambda_j^2 (\theta_j^\circ - \theta_j)^2}{\varepsilon^2} \leq \log(1 + 2\alpha^2)$ (similarity)
Note that $\lambda_j^2 a_j^2 \leq \varepsilon^2$ for all $j > \kappa_\star$. Hence,

$$\sum_{j\in\mathbb{N}} \lambda_j^2 \left(\theta_j^\circ - \theta_j\right)^2 = \xi^2 \varepsilon^4 \sum_{j\leq\kappa_\star} \lambda_j^2 \frac{L_j^2}{\lambda_j^4 \rho_\star^2} + \xi^2 \sum_{j>\kappa_\star} \lambda_j^2 a_j^4 \frac{L_j^2}{\rho_\star^2}$$
$$\leq \xi^2 \varepsilon^2 \sum_{j\leq\kappa_\star} \varepsilon^2 \frac{L_j^2}{\lambda_j^2 \rho_\star^2} + \xi^2 \varepsilon^2 \sum_{j>\kappa_\star} a_j^2 \frac{L_j^2}{\rho_\star^2}$$
$$\leq \xi^2 \varepsilon^2 \frac{\rho_\star^2}{\rho_\star^2} = \xi^2 \varepsilon^2 \leq \log(1 + 2\alpha^2) \varepsilon^2.$$

The conditions (1) and (2) guarantee that the constructed candidate sequence $\theta_\bullet$ is contained in the alternative. Condition (3) implies that $\chi^2(\mathbb{P}_{\theta^\circ_\bullet}, \mathbb{P}_{\theta_\bullet}) \leq 2\alpha^2$, which yields

$$\inf_{\Delta} \mathcal{R}\left(\Delta \mid \Theta^{\mathrm{R}}_{a_\bullet}, \underline{A}_\alpha \rho_\star\right) \geq 1 - \alpha$$

and, thus, completes the proof. □

2.4 Comparison to goodness-of-fit testing

In this section we illustrate the minimax radii of testing obtained in Proposition 2.2.1 and Proposition 2.3.1 and compare them to the radii obtained in the previous chapter.

Illustration 2.4.1. We calculate the order of the optimal dimension $\kappa_\star$ and the minimax radius of testing $\rho_\star$ for some specific combinations of the behaviour of $\lambda_\bullet$ (parameter p for ill-**p**osedness), $L_\bullet$ (parameter r for **R**iesz **R**epresentation of the linear functional) and $a_\bullet$ (parameter s for **s**moothness). We assume $s > \frac{1}{2}$ for the polynomially decaying case, which guarantees $a_\bullet \in \ell^2$.

Order of the optimal dimension $\kappa_\star$ and the minimax radius $\rho_\star$ for a linear functional L with $L_j = j^{-r}$				
a_j (smoothness)	λ_j (ill-posedness)	$\kappa_\star$	$\rho_\star$	
j^{-s}	j^{-p}	$\varepsilon^{-\frac{1}{s+p}}$	ε	$r > \frac{1}{2} + p$
			$\varepsilon\sqrt{\lvert\log\varepsilon\rvert}$	$r = \frac{1}{2} + p$
			$\varepsilon^{\frac{2s+2r-1}{2s+2p}}$	$r < \frac{1}{2} + p$
j^{-s}	e^{-jp}	$\lvert\log\varepsilon\rvert$	$\lvert\log\varepsilon\rvert^{\frac{1}{2}-(s+r)}$	
e^{-js}	j^{-p}	$\lvert\log\varepsilon\rvert$	ε	$r > \frac{1}{2} + p$
			$\varepsilon\sqrt{\log\lvert\log\varepsilon\rvert}$	$r = \frac{1}{2} + p$
			$\varepsilon\,\lvert\log\varepsilon\rvert^{\frac{1}{2}+(p-r)}$	$r < \frac{1}{2} + p$

Calculations for the risk bounds of Illustration 2.4.1.

1. **(mildly ill-posed – ordinary smooth)** The optimal $\kappa_\star$ is given by $\kappa_\star \sim \varepsilon^{-\frac{1}{s+p}}$. The variance term $\sqrt{\sum_{j\in[\![k]\!]} \frac{L_j^2}{\lambda_j^2}} = \sqrt{\sum_{j\in[\![k]\!]} j^{-2(r-p)}}$ behaves like a constant for $r > \frac{1}{2} + p$, like $\sqrt{\log k}$ for $r = \frac{1}{2} + p$ and like $k^{(p-r)+1/2}$ for $r < \frac{1}{2} + p$. The bias term $\sqrt{\sum_{j>k} L_j^2 a_j^2}$ is of order $k^{-(s+r)+1/2}$. Hence, in the case $r > \frac{1}{2} + p$ the rate is parametric. In the case $r = \frac{1}{2} + p$ it satisfies $\rho_\star \sim \varepsilon\sqrt{\log\kappa_\star} \sim \varepsilon\sqrt{\lvert\log\varepsilon\rvert}$. In the case $r < \frac{1}{2} + p$ it satisfies $\rho_\star \sim \kappa_\star^{-(s+r)+1/2} \sim \varepsilon^{\frac{2s+2r-1}{2s+2p}}$.

2. **(strongly ill-posed – ordinary smooth)** For the variance term $\sqrt{\sum_{j\in[\![k]\!]} \frac{L_j^2}{\lambda_j^2}}$ we have $\exp(kp) \gtrsim \sqrt{\sum_{j\in[\![k]\!]} \frac{L_j^2}{\lambda_j^2}} \gtrsim \exp(k(p-\delta))$ for any $\delta > 0$. The bias term $\sqrt{\sum_{j>k} L_j^2 a_j^2}$ is of order $k^{-(s+r)+1/2}$. The optimal dimension is of order $\kappa_\star \sim \lvert\log(\varepsilon)\rvert$, which yields a rate of order $\rho_\star \sim \lvert\log\varepsilon\rvert^{\frac{1}{2}-(s+r)}$.

3. **(mildly ill-posed – very smooth)** The variance term $\sqrt{\sum_{j\in[\![k]\!]} \frac{L_j^2}{\lambda_j^2}} = \sqrt{\sum_{j\in[\![k]\!]} j^{-2(r-p)}}$ behaves like a constant for $r > \frac{1}{2} + p$, like $\sqrt{\log k}$ for $r = \frac{1}{2} + p$ and like $k^{(p-r)+1/2}$ for $r < \frac{1}{2} + p$. The bias term $\sqrt{\sum_{j>k} L_j^2 a_j^2}$ satisfies $\exp(-ks) \gtrsim \sqrt{\sum_{j>k} L_j^2 a_j^2} \gtrsim \exp(-k(s+\delta))$ for any $\delta > 0$. The optimal dimension is of order $\kappa_\star \sim \lvert\log(\varepsilon)\rvert$. Hence, in the case $r > \frac{1}{2} + p$ the rate is parametric. In the case $r = \frac{1}{2} + p$ it satisfies $\rho_\star \sim \varepsilon\sqrt{\log\kappa_\star} \sim \varepsilon\sqrt{\log\lvert\log\varepsilon\rvert}$. In the case $r < \frac{1}{2} + p$ it satisfies $\rho_\star \sim \varepsilon\,\lvert\log\varepsilon\rvert^{\frac{1}{2}+(p-r)}$.

□

Note that if the linear functional $L_\bullet$ is smooth (fast decay of the coefficients, e.g. $L_j = e^{-j^r}$) such that it evens out the ill-posedness $\lambda_\bullet$ of the model, i.e. $\sum_{j\in\mathbb{N}} \frac{L_j^2}{\lambda_j^2} < \infty$, then variance term is uniformly bounded for all dimension parameters k. In this case we obtain a parametric rate $\rho_\star = \varepsilon$. Roughly speaking, the linear functional has a smoothing effect in both the bias and the variance part of the separation radius, which yields much smaller radii than the ones obtained in the classical goodness of fit-testing task (compare the rates in Illustration 1.2.5). A heuristic explanation for this phenomenon is the fact that it is naturally a much simpler task to only detect certain features $L(\theta_\bullet)$ of a sequence than to test the entire sequence $\theta_\bullet$. As an example consider the linear functional $L(\theta_\bullet) = \theta_1$, which extracts the first component of a sequence. Testing only one component of $\theta_\bullet$ can be done with a parametric rate, whereas testing the (infinite) sequence $\theta_\bullet$ yields a much slower (necessarily non-parametric) rate.

Let us explore the link between linear functional testing and goodness-of-fit testing in detail, i.e. we compare testing

$$H_0^{\mathrm{LF}} : L(\theta_\bullet) = L^\circ \qquad \text{against} \qquad H_1^{\mathrm{LF},\tilde{\rho}} : |L(\theta_\bullet) - L^\circ| \geq \tilde{\rho},\ \theta_\bullet - \theta_\bullet^\circ \in \Theta_{a_\bullet}^{\mathrm{R}}, \tag{LF}$$

with

$$H_0^{\mathrm{GoF}} : \theta_\bullet = \theta_\bullet^\circ \qquad \text{against} \qquad H_1^{\mathrm{GoF},\rho} : \|\theta_\bullet - \theta_\bullet^\circ\|_{\ell^2} \geq \rho,\ \theta_\bullet - \theta_\bullet^\circ \in \Theta_{a_\bullet}^{\mathrm{R}}. \tag{GoF}$$

More specifically, we aim to answer the question: how does an (optimal) test in one framework perform in the other? The next lemma shows that the null hypothesis of (LF) and the alternative of (GoF) intersect. Hence, a test cannot be of low level and powerful for both testing problems simultaneously.

Lemma 2.4.2 (Non-empty intersection of hypotheses). Let $\rho(\varepsilon) \searrow 0$ for $\varepsilon \searrow 0$. For ε small enough, there exists a sequence $\vartheta_\bullet \in \ell^2$ such that with $\rho = \rho(\varepsilon)$ the following two conditions are satisfied;

- $\vartheta_\bullet \in H_0^{\mathrm{LF}}$, i.e. $L(\vartheta_\bullet) = L^\circ$,
- $\vartheta_\bullet \in H_1^{\mathrm{GoF},\rho}$, i.e. $\|\vartheta_\bullet - \theta_\bullet^\circ\|_{\ell^2} \geq \rho$ and $\vartheta_\bullet - \theta_\bullet^\circ \in \Theta_{a_\bullet}^{\mathrm{R}}$.

Proof of Lemma 2.4.2. We distinguish two cases with respect to the linear functional.
First case: There exists an index $j_\circ$ such that $L_{j_\circ} = 0$. Then, let $\vartheta_\bullet = (\vartheta_j)_{j\in\mathbb{N}}$ be given by

$$\vartheta_j := \mathrm{R} a_{j_\circ} \mathbb{1}_{\{j=j_\circ\}} + \theta_j^\circ, \qquad j \in \mathbb{N}.$$

Then, $L(\vartheta_j) = L^\circ$, $\sum_{j\in\mathbb{N}} \left(\vartheta_j - \theta_j^\circ\right)^2 a_j^{-2} = \mathrm{R}^2$ and for ε small enough

$$\|\vartheta_\bullet - \theta_\bullet^\circ\|_{\ell^2}^2 = \mathrm{R}^2 a_{j_\circ}^2 \geq \rho^2(\varepsilon).$$

Hence, $\vartheta_\bullet \in H_0^{\mathrm{LF}} \cap H_1^{\mathrm{GoF},\rho}$.
Second case: We have $L_j \neq 0$ for all $j \in \mathbb{N}$. Define

$$\vartheta_1 := \frac{\mathrm{R}}{L_1} \cdot \frac{1}{\sqrt{\frac{1}{L_1^2 a_1^2} + \frac{1}{L_2^2 a_2^2}}} + \theta_1^\circ,$$
$$\vartheta_2 := -\frac{\mathrm{R}}{L_2} \cdot \frac{1}{\sqrt{\frac{1}{L_1^2 a_1^2} + \frac{1}{L_2^2 a_2^2}}} + \theta_2^\circ$$

and

$$\vartheta_j := \theta_j^\circ, \qquad j > 3.$$

Proceeding as above completes the proof. □

Let us now come back to answering our previous questions. Let Δ be a level-α-test for (LF), i.e. $\mathbb{P}_{\theta_\bullet}(\Delta = 1) \leq \alpha$ for all $\theta_\bullet$ with $L(\theta_\bullet) = L^\circ$. Naturally, we also have

$$\mathbb{P}_{\theta_\bullet^\circ}(\Delta = 1) \leq \sup_{\theta_\bullet \in H_0^{\mathrm{LF}}} \mathbb{P}_{\theta_\bullet}(\Delta = 1) \leq \alpha,$$

i.e. Δ is also a level-α-test for (GoF). However, due to *Lemma* 2.4.2 for a small enough noise level there exists $\vartheta_\bullet \in H_0^{\mathrm{LF}} \cap H_1^{\mathrm{GoF},\rho}$. Hence,

$$\sup_{\theta_\bullet \in H_1^{\mathrm{GoF}}} \mathbb{P}_{\theta_\bullet}(\Delta = 0) \geq \mathbb{P}_{\vartheta_\bullet}(\Delta = 0) \geq 1 - \alpha,$$

which implies that the power of Δ for the problem (GoF) cannot be larger than α.
Let Δ now be a $(1-\beta)$-powerful test for (GoF), i.e. $\mathbb{P}_{\theta_\bullet}(\Delta = 0) \leq \beta$ for all $\theta_\bullet \in H_1^{\mathrm{GoF},\rho}$. If L is an ℓ^2-functional, i.e. $L_\bullet \in \ell^2$, then $|L(\theta_\bullet - \theta_\bullet^\circ)| \leq \|L_\bullet\|_{\ell^2}^2 \|\theta_\bullet - \theta_\bullet^\circ\|_{\ell^2}^2$. Therefore,

$$|L(\theta_\bullet) - L^\circ| \geq \tilde{\rho} \qquad \Longrightarrow \qquad \|\theta_\bullet - \theta_\bullet^\circ\|_{\ell^2}^2 \geq \|L_\bullet\|_{\ell^2}^{-2} \tilde{\rho}.$$

Hence, with $\tilde{\rho} := \|L_\bullet\|_{\ell^2}^2 \rho^2$ we have

$$\sup_{\theta_\bullet \in H_1^{\mathrm{LF},\tilde{\rho}}} \mathbb{P}_{\theta_\bullet}(\Delta = 0) \leq \sup_{\theta_\bullet \in H_1^{\mathrm{GoF},\rho}} \mathbb{P}_{\theta_\bullet}(\Delta = 0) \leq \beta,$$

i.e. the type II error probability can be controlled. On the other hand, since Lemma 2.4.2 implies that there exists a $\vartheta_\bullet \in H_0^{\mathrm{LF}} \cap H_1^{\mathrm{GoF},\rho}$ we obtain for the type I error probability

$$\sup_{\theta_\bullet \in H_0^{\mathrm{LF}}} \mathbb{P}_{\theta_\bullet}(\Delta = 1) \geq \mathbb{P}_{\vartheta_\bullet}(\Delta = 1) \geq 1 - \beta.$$

Summarizing, a test intended for (LF) should not be used to test (GoF) and vice versa. This is a natural conclusion: a linear functional test is constructed to detect only a feature $L(\theta_\bullet)$ of the sequence of interest, which is a simpler task than to make inference on the entire sequence $\theta_\bullet$. Therefore, the two problems (GoF) and (LF) indeed require different testing strategies.

Remark 2.4.3 (Adaptation). *The test* (2.2.2) *relies on the knowledge of the regularity class* $\Theta_{a_\bullet}^{\mathrm{R}}$, *specifically on the sequence* $a_\bullet$ *and is, thus, non-adaptive. Let us briefly outline a possible strategy to obtain an adaptive test. By modifying the threshold* (2.2.1) *to only consist of a variance-type term it is possible to construct a (modified) test of the form* (2.2.2)*, which only depends on the regularity sequence* $a_\bullet$ *through the optimal choice of the dimension parameter* k*. In this situation the standard procedure to obtain an adaptive test is to aggregate the tests over various dimension parameters into a* max*-test, which rejects the null as soon as one of the tests does. Let us briefly discuss the effect of such a aggregation in the setting of this chapter. It is well-known (DasGupta [2008], Example 8.13) that for small* α *we approximately have* $\mathrm{q}_\alpha \approx \sqrt{2|\log \alpha|}$ *for the quantiles of a standard normal distribution. This is due to*

$$1 - \Psi(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{t^2}{2}} dt \leq \frac{1}{\sqrt{2\pi}} \int_x^\infty \frac{t}{x} e^{-\frac{t^2}{2}} dt = \frac{e^{-\frac{x^2}{2}}}{x\sqrt{2\pi}} = \frac{\phi(x)}{x}$$

where $\Psi(x)$ *and* $\phi(x)$ *denote the probability distribution function and the probability density function of a standard normal distribution, respectively. Hence, aggregation of the test* (2.2.2) *via the Bonferroni method over a finite collection* $\mathcal{K} \subseteq \mathbb{N}$ *of dimensions* k*, i.e. replacing* α *by* $\alpha/|\mathcal{K}|$ *yields an additional factor of* $\sqrt{\log|\mathcal{K}|}$ *in the testing radius. Since* $\mathcal{K}$ *is commonly chosen to be of cardinality* $|\log \varepsilon|$ *(by considering a geometric grid), the additional factor (i.e. a deterioration due to the aggregation) is of order* $\sqrt{\log|\log\varepsilon|}$ *(this translates to an* ***adaptive*** *factor of order* $(\log|\log\varepsilon|)^{1/4}$*), which we conjecture to be optimal (compare* Section 1.5 *in the previous chapter).* □

Appendix B

Auxiliary results

B.1 Calculations for the χ^2-divergence

Lemma B.1.1 (χ^2-divergence between two normals). Denote by $\mathbb{P}_{\mu_\bullet}$ respectively $\mathbb{P}_{\nu_\bullet}$ the probability measures associated with sequences of normal distributions with independent coordinates $\mathcal{N}(\mu_j, \varepsilon)$, $j \in \mathbb{N}$ respectively $\mathcal{N}(\nu_j, \varepsilon)$, $j \in \mathbb{N}$ and $\varepsilon > 0$. Then, the χ^2-divergence satisfies

$$\chi^2(\mathbb{P}_{\mu_\bullet}, \mathbb{P}_{\nu_\bullet}) = \exp\left(\sum_{j\in\mathbb{N}} \frac{(\nu_j - \mu_j)^2}{\varepsilon^2}\right) - 1$$

Proof of Lemma B.1.1. We recall the definition of the χ^2-divergence between two measures $\mathbb{P}, \mathbb{Q}$ (cp. Tsybakov [2009], Section 2.4)

$$\chi^2(\mathbb{P}, \mathbb{Q}) = \begin{cases} \int \left(\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}} - 1\right)^2 \mathrm{d}\mathbb{Q} & \text{if } \mathbb{P} \ll \mathbb{Q} \\ \infty & \text{otherwise .} \end{cases}$$

Since in our case both $\mathbb{P}_{\nu_\bullet}$ and $\mathbb{P}_{\mu_\bullet}$ are normal distributions we have $\mathbb{P}_{\mu_\bullet} \ll \mathbb{P}_{\nu_\bullet}$. Furthermore,

$$\chi^2(\mathbb{P}, \mathbb{Q}) = \int \left(\left(\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}\right)^2 - 2\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}} + 1\right) \mathrm{d}\mathbb{Q} = \int \left(\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}\right)^2 \mathrm{d}\mathbb{Q} - 1 = \mathbb{E}_{Z\sim\mathbb{Q}} \left(\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}(Z)\right)^2 - 1.$$

Let us now first determine the likelihood ratio

$$\frac{\mathrm{d}\mathbb{P}_{\mu_\bullet}}{\mathrm{d}\mathbb{P}_{\nu_\bullet}}(y_\bullet) = \prod_{j\in\mathbb{N}} \exp\left(\frac{(y_j - \nu_j)^2 - (y_j - \mu_j)^2}{2\varepsilon^2}\right) = \prod_{j\in\mathbb{N}} \exp\left(\frac{2y_j(\mu_j - \nu_j) + (\nu_j^2 - \mu_j^2)}{2\varepsilon^2}\right).$$

Let $Y_\bullet$ be a random variable with distribution $\mathbb{P}_{\nu_\bullet}$. We obtain

$$\mathbb{E}_{\nu_\bullet} \left(\frac{\mathrm{d}\mathbb{P}_{\mu_\bullet}}{\mathrm{d}\mathbb{P}_{\nu_\bullet}}(Y_\bullet)\right)^2$$
$$= \prod_{j\in\mathbb{N}} \int \frac{1}{\sqrt{2\pi\varepsilon^2}} \exp\left(\frac{4y_j(\mu_j - \nu_j) + 2(\nu_j^2 - \mu_j^2)}{2\varepsilon^2}\right) \exp\left(\frac{-(y_j - \nu_j)^2}{2\varepsilon^2}\right) \mathrm{d}y_j.$$

Inside of the exponential function in the integral we carry out a square addition

$$4y_j(\mu_j - \nu_j) - (y_j - \nu_j)^2 = -(y_j - (2\mu_j - \nu_j))^2 + (2\mu_j - \nu_j)^2 - \nu_j^2,$$

which yields

$$
\begin{aligned}
&\mathbb{E}_{\nu_\bullet}\left(\frac{\mathrm{d}\mathbb{P}_{\mu_\bullet}}{\mathrm{d}\mathbb{P}_{\nu_\bullet}}(Y_\bullet)\right)^2 \\
&= \prod_{j\in\mathbb{N}} \exp\left(\frac{2(\nu_j^2-\mu_j^2)+(2\mu_j-\nu_j)^2-\nu_j^2}{2\varepsilon^2}\right) \underbrace{\int \frac{1}{\sqrt{2\pi\varepsilon^2}} \exp\left(\frac{-(y_j-(2\mu_j-\nu_j))^2}{2\varepsilon^2}\right) \mathrm{d}y_j}_{=1} \\
&= \prod_{j\in\mathbb{N}} \exp\left(\frac{(\nu_j-\mu_j)^2}{\varepsilon^2}\right) = \exp\left(\sum_{j\in\mathbb{N}} \frac{(\nu_j-\mu_j)^2}{\varepsilon^2}\right),
\end{aligned}
$$

which completes the proof. □

Part II

Circular convolution

Circular convolution

We consider a circular convolution model, where a random variable that takes values on the circle is observed contaminated by an additive error. Identifying the circle with the unit interval $[0, 1)$, the observable random variable is given by

$$Y := X + \varepsilon - \lfloor X + \varepsilon \rfloor = X + \varepsilon \mod 1,$$

where X and ε are independent random variables supported on the interval $[0, 1)$ and $\lfloor \cdot \rfloor$ denotes the floor-function. The next proposition characterises the density of the random variable Y in terms of the densities of X and ε, i.e. the density of Y is the **circular convolution** of the densities of X and ε.

Proposition (Convolution density). Let $X \sim f$ and $\varepsilon \sim \varphi$ be independent random variables on $[0, 1)$. The random variable $Y := X + \varepsilon - \lfloor X + \varepsilon \rfloor$ has density $g = f \circledast \varphi$ with

$$g(y) := (f \circledast \varphi)(y) := \int_{[0,1)} f((y - s) \mod 1)\varphi(s)\mathrm{d}s, \qquad y \in [0, 1).$$

Proof. Let $y \in [0, 1)$. By independence of X and ε we have

$$F^Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(X + \varepsilon \mod 1 \leq y)$$
$$= \int_{[0,1)} \int_{[0,1)} f(x)\varphi(s)\mathbb{1}_{\{x+s \leq y \mod 1\}}\mathrm{d}x\mathrm{d}s,$$

where we introduce the change of variable, $t = x + s \mod 1$, $\mathrm{d}x = \mathrm{d}t$ and obtain

$$F^Y(y) = \int_{[0,1)} \int_{[0,1)} f(t - s \mod 1)\varphi(s)\mathbb{1}_{\{t \leq y \mod 1\}}\mathrm{d}s\mathrm{d}t$$
$$= \int_{[0,y)} \int_{[0,1)} f(t - s \mod 1)\varphi(s)\mathrm{d}s\mathrm{d}t.$$

Taking the derivative yields the desired result

$$g(t) = \frac{\mathrm{d}}{\mathrm{d}y} F^Y(y) \mid_{y=t} = \int_{[0,1)} f(t - s \mod 1)\varphi(s)\mathrm{d}s.$$

□

Related literature. Circular data, also called wrapped (around the circumference of the unit circle), spherical or directional, appears in various applications. For an in-depth review of many examples for circular data we refer the reader to Mardia [1972], Fisher [1995] and Mardia and Jupp [2009]. Let us only briefly mention two popular fields of application. Circular models are used for data with a temporal or periodic structure, where the circle is identified e.g. with a clock face (cp. Gill and Hangartner [2010]). Moreover, also directional data can be represented by a circular model by identifying the circle with a compass rose. Kerkyacharian et al. [2011] and Lacour and Ngoc [2014], for instance, investigate a circular model with multiplicative error. Nonparametric estimation in the additive error model has amongst others been considered in Efromovich [1997], Comte and Taupin [2003] and Johannes and Schwarz [2013].

Some examples of circular densities. Figure 2.1 displays several typical densities on the circle, Figure 2.2 plots the same densities on $[0, 1)$ for comparison. More details and many more examples can be found in chapter 3 of the textbook Mardia and Jupp [2009].

1. **(von Mises distribution)** This is the analogue of the Gaussian distribution on the circle, hence, also known as the *circular normal distribution.* Denote by $I_\circ(\kappa)$ the modified Bessel-function of order 0, then the density of a von Mises $\mathrm{vM}(\mu, \kappa)$-distribution with location parameter μ and measure of concentration $\kappa > 0$ is given by

   $$f(x) = \frac{1}{I_\circ(\kappa)} \exp(\kappa \cos(2\pi(x - \mu)))$$

   We plot the density for $\mu = 1/2$ and $\kappa = 3$.

2. **(Uniform distribution)** The density of a uniform $U[0, 1)$ distribution is given by

   $$f(x) = \mathbb{1}_{[0,1)}.$$

   For $\kappa \longrightarrow 0$ the von Mises distribution approaches the uniform distribution on the circle.

3. **(Cardioid Distribution)** For a parameter $|\rho| < \frac{1}{2}$ and a location parameter μ the cardioid distribution $\mathrm{C}(\mu, \rho)$ has density

   $$f(x) = 1 + 2\rho \cos(2\pi(x - \mu)).$$

   For small κ the von Mises distribution approximates the Cardioid distribution since $\exp(\kappa) \approx 1 + \kappa$. We plot the density with $\mu = 1/2$ and $\rho = 0.25$.

4. **(Triangular Distribution)** The density of a triangular distribution is given by

   $$f(x) = \begin{cases} x & \text{for } x < 1/2, \\ 1 - x & \text{for } x > 1/2. \end{cases}$$

5. **(Wrapped Normal distribution)** A distribution on the circle can be obtained by wrapping a distribution given on the real line around the circumference of the unit circle. Let $R \sim \mathcal{N}(\mu, \sigma^2)$ be a normally distributed random variable on $\mathbb{R}$, then

   $$X := R - \lfloor R \rfloor = R \mod 1$$

   is a random variable on the unit circle and we denote its distribution by $\mathrm{WN}(\mu, \sigma^2)$. If ξ is the density of R, then the density of X is given by

   $$f(x) = \sum_{k \in \mathbb{Z}} \xi(2\pi(x + k)).$$

   We plot the wrapped normal distribution for $\mu = 1/2$ and $\sigma = 1$.

6. **(Wrapped Cauchy distribution)** In the same way as in the example of the Normal distribution, we can wrap a Cauchy distribution $\mathrm{Cauchy}(s, t)$ around the unit circle. Its density is plotted for $s = 1/2$ and $t = e$.

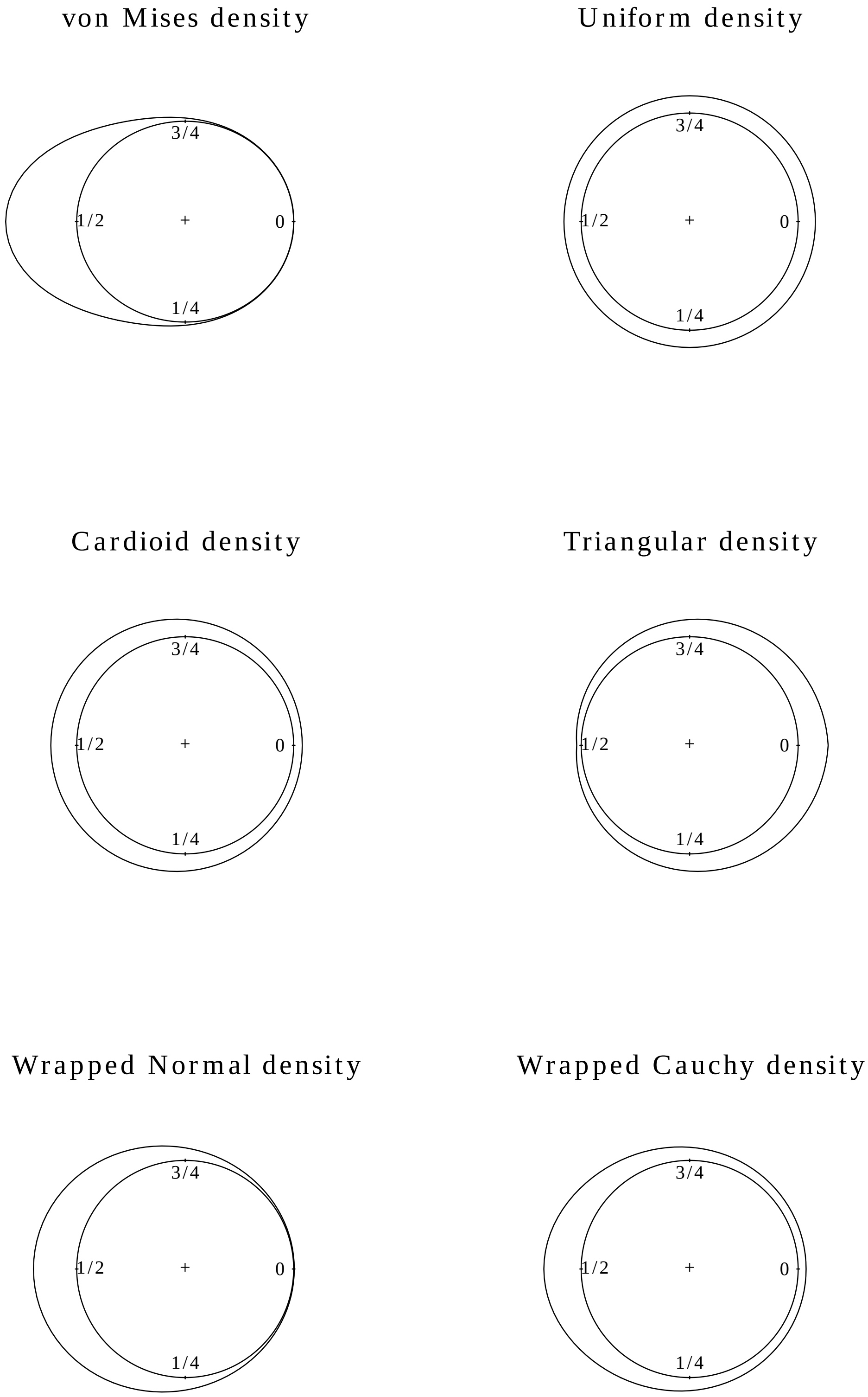


Figure 2.1: Typical densities on the unit circle.

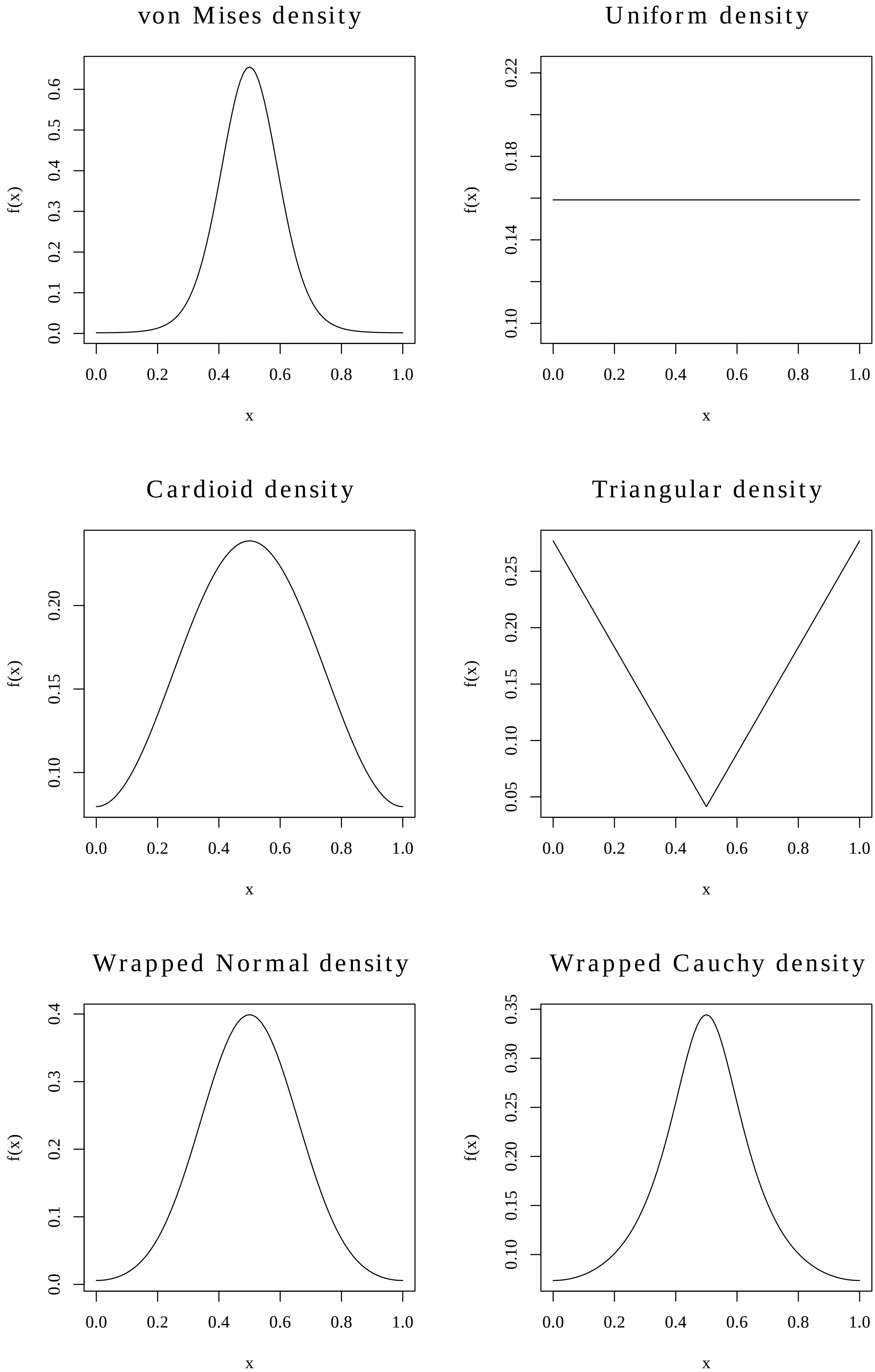


Figure 2.2: Typical densities on the unit circle plotted on $[0, 1)$.

Some examples of circular data.

1. **(temporal data)** Figure 2.3 presents the arrival times of 254 patients at an intensive care unit, the data was collected over a period of 12 months. The data points are taken from Fisher [1995], p. 239 and were originally published in Cox and Lewis [1966], p. 254-255.

2. **(directional data)** Figure 2.4 shows two different ways of representing circular data in a diagram. We are given the orientation of 76 turtles after laying eggs (data taken from Table I.5 in Mardia and Jupp [2009]). On the left the data points are represented as a rose diagram (the circular analogue of a histogram), on the right the points are plotted around the circle and their estimated density is presented in green. For comparison we also plot the density of a mixture of von Mises distributions, one with weight 0.8 and location parameter 1/6 (roughly the direction of the sea) and a second with weight 0.2 and location parameter 4/6 (direction away from the sea), both with measure of concentration $\kappa = 3$.

3. **(temporal data)** Figure 2.5 shows the estimated density of the times of birth in the US (2018) (red line) with 3801534 data points obtained from `https://www.cdc.gov/nchs/data_access/vitalstatsonline.htmm` plotted around a 24-hour clock face. For comparison we plot the density of a uniform distribution. An interesting question could be: Is the birth time uniformly distributed around the 24h-clockface? Our goodness-of-fit tests proposed in Chapter 3 and Chapter 4 are based on an estimation of the green area.

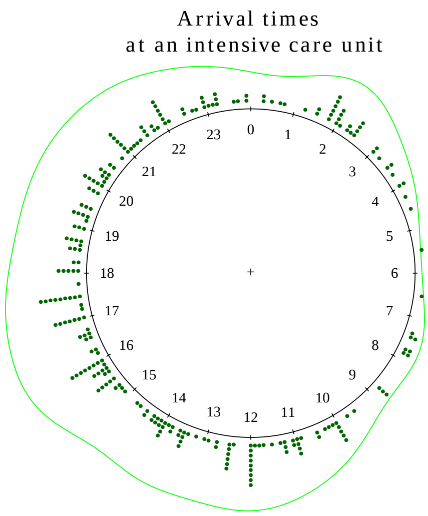


Figure 2.3: Arrival times at an intensive care unit plotted around a 24-hour clock face (green dots) and their estimated density (green line).

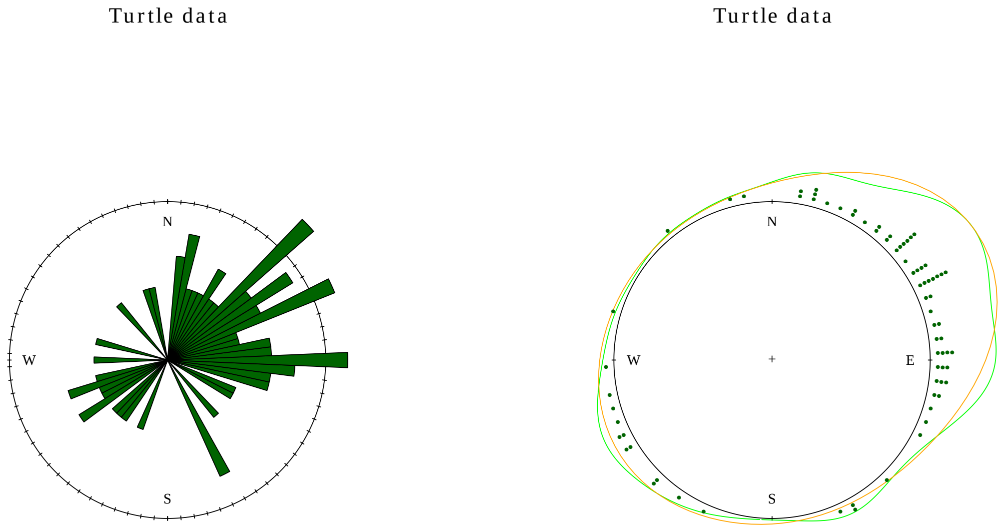


Figure 2.4: Orientation in which 76 turtles leave their nest after laying eggs, represented as a rose diagram (left) and as points on the circle (right). The green line is the estimated density, for comparison we plot the density of a mixture of two von Mises distributions (orange line).

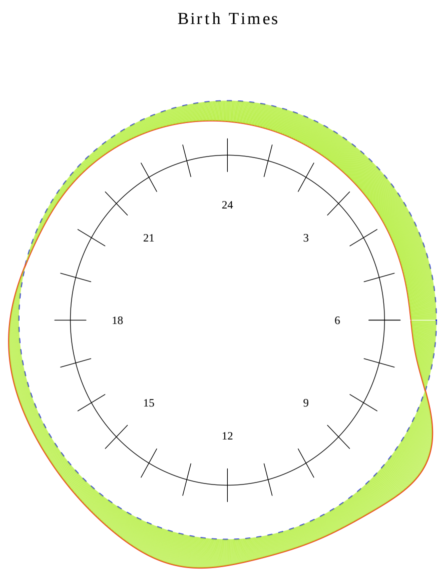


Figure 2.5: Estimated density of the times of birth in the US (2018) (red line) and its distance (green) to a uniform distribution (blue dashed line).

Fourier coefficients property. We denote by $\mathscr{L}^2 = \mathscr{L}^2[0,1)$ the space of square-integrable complex-valued functions on $[0,1)$ equipped with its usual inner product. The methodology of the following chapters relies on the expansion of the respective functions in the exponential basis $\boldsymbol{e_j}, j \in \mathbb{Z}$ of $\mathscr{L}^2$ with $\boldsymbol{e_j}(x) = \exp(2\pi i j x)$ for $x \in [0,1)$. The next proposition provides a useful property for circularly convoluted densities in terms of their Fourier coefficients. The proof is similar to the well-known proof for Fourier transforms and convolution on $\mathbb{R}$ and we state it here for completeness (cp. Appendix A in Meister [2009], Lemma A.1)

Proposition (Circular convolution theorem). Let $f, \varphi \in \mathscr{L}^2$ with Fourier coefficients $f_\bullet = (f_j)_{j\in\mathbb{Z}} = (\langle f, \boldsymbol{e_j}\rangle_{\mathscr{L}^2})_{j\in\mathbb{Z}}$ and $\varphi_\bullet = (\varphi_j)_{j\in\mathbb{Z}} = (\langle \varphi, \boldsymbol{e_j}\rangle_{\mathscr{L}^2})$. We have $g = f \circledast \varphi$ if and only if $g_j = f_j \cdot \varphi_j$ for all $j \in \mathbb{Z}$.

Proof. Let $g = f \circledast \varphi$, then for $j \in \mathbb{Z}$ we have

$$\begin{aligned} f_j \cdot \varphi_j &= \int_{[0,1)} f(x)\boldsymbol{e_j}(-x)\mathrm{d}x \int_{[0,1)} \varphi(y)\boldsymbol{e_j}(-y)\mathrm{d}y \\ &= \int_{[0,1)} \int_{[0,1)} f(x)\boldsymbol{e_j}(-x)\varphi(y)\boldsymbol{e_j}(-y)\mathrm{d}y\mathrm{d}x, \end{aligned}$$

where we introduce the change of variable, $z = x + y \mod 1, \mathrm{d}x = \mathrm{d}z$ and obtain

$$\begin{aligned} f_j \cdot \varphi_j &= \int_{[0,1)} \int_{[0,1)} f(z - y \mod 1)\varphi(y)\boldsymbol{e_j}(-z)\mathrm{d}z\mathrm{d}y \\ &= \int_{[0,1)} (f \circledast \varphi)(z)\boldsymbol{e_j}(-z)\mathrm{d}z = \int_{[0,1)} g(z)\boldsymbol{e_j}(-z) = g_j. \end{aligned}$$

Now let $g_j = f_j \cdot \varphi_j$ for all $j \in \mathbb{Z}$, then by the Riesz-Fischer Theorem we have the representation (with equality in $\mathscr{L}^2$)

$$g(y) = \sum_{j\in\mathbb{Z}} g_j \boldsymbol{e_j}(y) = \sum_{j\in\mathbb{Z}} f_j\varphi_j\boldsymbol{e_j}(y) = \sum_{j\in\mathbb{Z}} \varphi_j \int_{[0,1)} f(x)\boldsymbol{e_j}(-x)\boldsymbol{e_j}(y)\mathrm{d}x,$$

where we introduce the change of variable, $x = y - s \mod 1, \mathrm{d}x = \mathrm{d}s$ and exploit the periodicity of $\boldsymbol{e_j}$, $j \in \mathbb{Z}$

$$\begin{aligned} g(y) &= \sum_{j\in\mathbb{Z}} \varphi_j \int_{[0,1)} f(y - s \mod 1)\boldsymbol{e_j}(-y + s)\boldsymbol{e_j}(y)\mathrm{d}s \\ &= \sum_{j\in\mathbb{Z}} \varphi_j \int_{[0,1)} f(y - s \mod 1)\boldsymbol{e_j}(s)\mathrm{d}s \\ &= \int_{[0,1)} f(y - s \mod 1) \sum_{j\in\mathbb{Z}} \varphi_j\boldsymbol{e_j}(s)\mathrm{d}s \\ &= \int_{[0,1)} f(y - s \mod 1)\varphi(s)\mathrm{d}s = (f \circledast \varphi)(y). \end{aligned}$$

□

Notation

In contrast to the previous part of this thesis, from here on we also consider complex-valued sequences and complex-valued functions. In this second part we denote

$$\begin{aligned} \ell^2 &:= \ell^2(\mathbb{Z}) := \left\{ x_\bullet \in \mathbb{C}^\mathbb{Z} : \sum_{j\in\mathbb{N}} x_j^2 < \infty \right\}, \\ \ell^\infty &:= \ell^\infty(\mathbb{Z}) := \left\{ x_\bullet \in \mathbb{C}^\mathbb{Z} : \sup_{j\in\mathbb{Z}} |x_j| < \infty \right\}. \end{aligned}$$

The space $\ell^2 := \ell^2(\mathbb{Z})$ equipped with $\langle x_\bullet, y_\bullet \rangle_{\ell^2} := \sum_{j\in\mathbb{N}} x_j \overline{y_j}$, $\|x_\bullet\|_{\ell^2}^2 := \sum_{j\in\mathbb{N}} |x_j|^2$ is a Hilbert space of square summable complex-valued sequences, ℓ^∞ equipped with $\|x_\bullet\|_{\ell^\infty} = \sup_{j\in\mathbb{Z}} |x_j|$ is a Banach space of bounded sequences.

By $\mathscr{L}^2 := \mathscr{L}^2[0,1)$ we denote in this part the Hilbert space of **complex**-valued square integrable functions defined on the half-open unit interval $[0,1)$ equipped with the inner product $\langle f, g \rangle_{\mathscr{L}^2} = \int_0^1 f(x)\overline{g(x)}\mathrm{d}x$.

Chapter 3

Minimax testing and quadratic functional estimation for circular convolution

In the circular convolution model we aim to infer on the density of a circular random variable using observations contaminated by an additive measurement error. We highlight the interplay of the two problems: optimal testing and quadratic functional estimation. Under general regularity assumptions we determine an upper bound for the minimax risk of estimation for the quadratic functional. The upper bound consists of two terms, one that mimics a classical bias^2-variance trade-off and a second that causes the typical elbow effect in quadratic functional estimation. Using a minimax optimal estimator of the quadratic functional as a test statistic, we derive an upper bound for the non-asymptotic minimax radius of testing for non-parametric alternatives. Interestingly, the term causing the elbow effect in the estimation case vanishes in the radius of testing. We provide a matching lower bound for the testing problem. By showing that any lower bound for the testing problem also yields a lower bound for the quadratic functional estimation problem, we obtain a lower bound for the risk of estimation. Lastly, we prove a matching lower bound for the term causing the elbow effect. Therefore, we establish both the minimax risk of estimation and the minimax radius of testing.

3.1 Introduction

The statistical model. In this section we consider minimax testing and quadratic functional estimation in a circular convolution model. We observe a random variable given by

$$Y := X + \varepsilon - \lfloor X + \varepsilon \rfloor,$$

where X and ε are independent random variables on $[0, 1)$ with densities f and φ, respectively. The density of the observable random variable Y satisfies $g = f \circledast \varphi$, where $\circledast$ denotes the circular convolution. The model is introduced and motivated in detail in the section above.

3.1.1 Quadratic functional estimation

Denote by $\mathcal{D}$ the subset of real probability densities in $\mathscr{L}^2 := \mathscr{L}^2[0, 1)$, the Hilbert space of square-integrable complex-valued functions on $[0, 1)$ equipped with its usual norm $\|\cdot\|_{\mathscr{L}^2}$. Since we are interested in the estimation of the quadratic functional $\mathrm{q}^2(f) := \|f\|^2_{\mathscr{L}^2}$ of a density f, we assume throughout this paper that both f and φ (and, hence, g) belong to $\mathcal{D}$. We also want to compare f to the prescribed density $f^\circ = \mathbb{1}_{[0,1)}$ of a uniform distribution by estimating their

$\mathscr{L}^2$-distance $\mathrm{q}^2(f-f^\circ) = \|f - f^\circ\|^2_{\mathscr{L}^2}$. Since $\mathrm{q}^2(f-f^\circ) = \mathrm{q}^2(f)-1$ these problems are equivalent and we focus on the estimation of $\mathrm{q}^2(f - f^\circ)$. Let $\{Y_k\}_{j=1}^n$ be a sample of n independent and identically distributed observations with density g, i.e. the observations are given by

$$Y_k \overset{\text{iid}}{\sim} g = f \circledast \varphi, \qquad k \in [\![n]\!]. \tag{3.1.1}$$

Denote by $\mathbb{P}_f$ and $\mathbb{E}_f$ the probability distribution and the expectation associated with the data (3.1.1). For a non-parametric class $\mathcal{E}$ we measure the accuracy of an estimator $\hat{\mathrm{q}}^2$, i.e. a measurable function $\hat{\mathrm{q}}^2 : \mathbb{R}^n \longrightarrow \mathbb{R}$ by its **maximal risk**

$$r^2(\hat{\mathrm{q}}^2, \mathcal{E}) := \sup_{f-f^\circ \in \mathcal{E}} \mathbb{E}_f \left(\hat{\mathrm{q}}^2 - \mathrm{q}^2(f - f^\circ)\right)$$

and compare its performance to the **minimax risk of estimation**

$$r^2(\mathcal{E}) := \inf_{\hat{\mathrm{q}}^2} r^2(\hat{\mathrm{q}}^2, \mathcal{E}),$$

where the infimum is taken over all possible estimators. An estimator $\hat{\mathrm{q}}^2$ is called **minimax optimal** for the class $\mathcal{E}$ if its maximal risk is bounded by the minimax risk $r^2(\mathcal{E})$ up to a constant. Note that with the non-parametric class $\mathcal{E}$ we put restrictions on the difference $f - f^\circ$ instead of on f directly, this makes it easier to compare the quadratic functional estimation problem with the testing problem, which we state below.

Related literature. Quadratic functional estimation in direct models has received much attention in the literature, let us only mention a few references. Bickel and Ritov [1988] and Birgé and Massart [1995] establish minimax rates for the estimation of functionals of a density, where they discover a typical phenomenon in quadratic functional estimation: the so-called **elbow effect**, which also appears in our results. It refers to a sudden change in the behaviour of the rates as soon as the smoothness/regularity parameter crosses a critical threshold. In a Gaussian sequence space model, which is closely related to our model, for instance, Laurent and Massart [2000] and Laurent [2005] consider adaptive quadratic functional estimation via model selection, Cai and Low [2005a] and Cai and Low [2006] derive minimax optimal estimators under Besov-type regularity assumptions. Collier et al. [2017] consider sparsity constraints. Quadratic functional estimation in an inverse Gaussian sequence space model is treated by Butucea and Meziani [2011] (known operator) and Kroll [2019a] (partially unknown operator). For quadratic functional estimation for deconvolution on the real line we refer to Butucea [2007] and Chesneau [2011].

3.1.2 The testing task

Based on the observations (3.1.1) we test the null hypothesis $\{f = f^\circ\}$ against the alternative $\{f \neq f^\circ\}$. To make the null hypothesis and the alternative distinguishable, we separate them in the $\mathscr{L}^2$-norm. For a separation radius $\rho \in \mathbb{R}_+$ let us define the set $\mathscr{L}^2_\rho := \{\xi \in \mathscr{L}^2 : \|\xi\|_{\mathscr{L}^2} \geq \rho\}$, which is called the **energy condition**. For a nonparametric class of functions $\mathcal{E}$, called the **regularity condition**, the testing problem can be written as

$$H_0 : f = f^\circ \qquad \text{against} \qquad H_1^\rho : f - f^\circ \in \mathscr{L}^2_\rho \cap \mathcal{E}, f \in \mathcal{D}. \tag{3.1.2}$$

We measure the accuracy of a test Δ, i.e. a measurable function $\Delta : \mathbb{R}^{\mathbb{N}} \longrightarrow \{0, 1\}$, by its maximal risk defined as the sum of the type I error probability and the maximal type II error probability over the ρ-separated alternative

$$\mathcal{R}\left(\Delta \mid \mathcal{E}, \rho\right) := \mathbb{P}_{f^\circ}(\Delta = 1) + \sup_{\substack{f-f^\circ \in \mathscr{L}^2_\rho \cap \mathcal{E} \\ f \in \mathcal{D}}} \mathbb{P}_f(\Delta = 0).$$

We aim to answer the question how far the null and the alternative need to be separated to be statistically distinguishable. A value $\rho^2(\Delta, \mathcal{E}) = \rho^2(\{\Delta_\alpha\}_{\alpha\in(0,1)}, \mathcal{E})$ is called **radius of testing** for the family of tests $\{\Delta_\alpha\}_{\alpha\in(0,1)}$ over the alternative $\mathcal{E}$, if for all $\alpha \in (0,1)$ there exist constants $\underline{A}_\alpha, \overline{A}_\alpha \in \mathbb{R}_+$ such that

(i) for all $A \geq \overline{A}_\alpha$ we have $\mathcal{R}\left(\Delta_\alpha \mid \mathcal{E}, A\rho(\Delta, \mathcal{E})\right) \leq \alpha,$ (upper bound)

(ii) for all $A \leq \underline{A}_\alpha$ we have $\mathcal{R}\left(\Delta_\alpha \mid \mathcal{E}, A\rho(\Delta, \mathcal{E})\right) \geq 1 - \alpha.$ (lower bound)

The difficulty of the testing problem can be characterised by the **minimax risk**

$$\mathcal{R}\left(\mathcal{E}, \rho\right) := \inf_{\Delta} \mathcal{R}\left(\Delta \mid \mathcal{E}, \rho\right)$$

where the infimum is taken over all possible tests. The value $\rho^2(\mathcal{E})$ is called **minimax radius of testing** if for all $\alpha \in (0,1)$ there exist constants $\underline{A}_\alpha, \overline{A}_\alpha \in \mathbb{R}_+$ such that

(i) for all $A \geq \overline{A}_\alpha$ we have $\mathcal{R}\left(\mathcal{E}, A\rho(\mathcal{E})\right) \leq \alpha,$ (upper bound)

(ii) for all $A \leq \underline{A}_\alpha$ we have $\mathcal{R}\left(\mathcal{E}, A\rho(\mathcal{E})\right) \geq 1 - \alpha.$ (lower bound)

If $\rho^2(\mathcal{E})$ is a radius of testing for the family of tests $\{\Delta_\alpha\}_{\alpha\in(0,1)}$, then it is called **minimax optimal**.

Related literature. Concerning minimax testing in convolution models we refer e.g. to Butucea [2007], Butucea et al. [2009] and Loubes and Marteau [2014], all three consider convolution on the real line. The connection between quadratic functional estimation and testing has for example been studied in Collier et al. [2017] (in a direct Gaussian sequence space model under sparsity), Kroll [2019a] (in a indirect Gaussian sequence space model under regularity constraints) and Butucea [2007] (in a convolution model on the real line).

3.1.3 Methodology

We characterise both the minimax risk and the minimax radius in terms of the sample size n, the parameters of the regularity class $\mathcal{E}$ and the error density φ. Our approach heavily depends on the properties of the Hilbert space $\mathscr{L}^2 := \mathscr{L}^2[0,1)$ equipped with its usual inner product $\langle \cdot, \cdot \rangle := \langle \cdot, \cdot \rangle_{\mathscr{L}^2}$ given by

$$\langle \xi, \psi \rangle := \int_{[0,1)} \xi(x)\overline{\psi(x)}\mathrm{d}x \qquad \text{for } \xi, \psi \in \mathscr{L}^2,$$

where $\overline{\psi(x)}$ denotes the complex conjugate of $\psi(x)$. Given the exponential basis $e_j, j \in \mathbb{Z}$ of $\mathscr{L}^2$ with $e_j(x) = \exp(2\pi i j x)$ for $x \in [0,1)$, we denote the Fourier coefficients of a function $f \in \mathscr{L}^2$ by $f_j = \langle f, e_j \rangle$, $j \in \mathbb{Z}$. This leads to the discrete Fourier series expansion

$$f = \sum_{j\in\mathbb{Z}} f_j e_j, \tag{3.1.3}$$

where equality holds in $\mathscr{L}^2$. The non-parametric class of functions $\mathcal{E}$ is formulated in terms of the Fourier coefficients and characterises the regularity of the function. Let $\mathrm{R} > 0$ and let $a_\bullet = (a_j)_{j\in\mathbb{N}} \in \mathbb{R}^\mathbb{N}$ be a strictly positive, monotonically non-increasing sequence. We assume that the density of interest f (resp. $f - f^\circ$) belongs to the $\mathscr{L}^2$-ellipsoid

$$\mathcal{E}_{a_\bullet}^{\mathrm{R}} := \left\{ \xi \in \mathscr{L}^2 : 2\sum_{j\in\mathbb{N}} a_j^{-2} \left|\xi_j\right|^2 \leq \mathrm{R}^2 \right\}. \tag{3.1.4}$$

We point out that in the case $f^\circ = \mathbb{1}_{[0,1)}$ the conditions $f \in \mathcal{E}^{\mathrm{R}}_{a_\bullet}$ and $f - f^\circ \in \mathcal{E}^{\mathrm{R}}_{a_\bullet}$ are equivalent. Moreover, note that for real-valued densities, the condition $f \in \mathcal{E}^{\mathrm{R}}_{a_\bullet}$ imposes conditions on all coefficients $f_j, j \in \mathbb{Z}$, since $|f_j|^2 = |f_{-j}|^2$, $j \in \mathbb{N}$ for all real-valued functions and, additionally $f_0 = 1$ for all densities. The definition (3.1.4) is general enough to cover classes of ordinary and super smooth densities. Expanding both f and f° in the exponential basis as in (3.1.3) and applying Parseval's Theorem yields a representation of the quadratic functional $\mathrm{q}^2(f - f^\circ) = \|f - f^\circ\|^2_{\mathscr{L}^2}$ in their Fourier coefficients $\mathrm{q}^2(f - f^\circ) = \sum_{j\in\mathbb{Z}} \left|f_j - f^\circ_j\right|^2 = 2\sum_{j\in\mathbb{N}} \left|f_j - f^\circ_j\right|^2$. In particular, for $f^\circ = \mathbb{1}_{[0,1)}$, we have $\mathrm{q}^2(f - f^\circ) = 2\sum_{j\in\mathbb{N}} |f_j|^2$. Moreover, by the circular convolution theorem we have $g = f \circledast \varphi$ if and only if the Fourier coefficients satisfy $g_j = f_j \cdot \varphi_j$ for all $j \in \mathbb{Z}$. Here and subsequently we assume that the Fourier coefficients of the noise density φ are non-vanishing everywhere, i.e. $|\varphi_j| > 0$ for all $j \in \mathbb{Z}$. The quadratic functional can then be expressed as

$$\mathrm{q}^2(f - f^\circ) = \sum_{j\in\mathbb{Z}} \frac{\left|g_j - \varphi_j f^\circ_j\right|^2}{|\varphi_j|^2}, \tag{3.1.5}$$

which simplifies to $\mathrm{q}^2(f - f^\circ) = 2\sum_{j\in\mathbb{N}} \frac{|g_j|^2}{|\varphi_j|^2}$ in the case of a uniform density $f^\circ = \mathbb{1}_{[0,1)}$. The only unknown quantities in (3.1.5) are the Fourier coefficients g_j, $j \in \mathbb{Z}$ of g, which can easily be estimated. Since for $j \in \mathbb{Z}$, $g_j = \langle g, e_j\rangle = \mathbb{E}_f e_j(-Y_1)$, a natural estimator is given by replacing the expectation with the empirical counterpart $\hat{g}_j := \frac{1}{n}\sum_{k=1}^n e_j(-Y_k)$. Inserting these estimators into the quadratic functional, however, generates a bias in every component. Since $|\hat{g}_j|^2 - \frac{1-|\hat{g}_j|^2}{n-1}$ is an unbiased estimator of the numerator $|g_j|^2$, for $j \in \mathbb{N}$, for each $k \in \mathbb{N}$ we consider the estimator

$$\hat{\mathrm{q}}^2_k := 2 \sum_{|j|\in[\![k]\!]} |\varphi_j|^{-2} \left\{ |\hat{g}_j|^2 - \frac{1 - |\hat{g}_j|^2}{n-1} \right\}, \tag{3.1.6}$$

which is an unbiased estimator of the truncated quadratic functional $\mathrm{q}^2_k(f - \mathbb{1}_{[0,1)}) := 2\sum_{j=1}^k \frac{|g_j|^2}{|\varphi_j|^2}$. Here and subsequently, we only consider the case $f^\circ = \mathbb{1}_{[0,1)}$. Using $\hat{\mathrm{q}}^2_k$ as an estimator of the distance $\|f - f^\circ\|^2_{\mathscr{L}^2}$ to the null hypothesis, we construct a test that, roughly speaking, compares the estimator to a multiple of its standard deviation. Precisely, for $k \in \mathbb{N}, \alpha \in (0,1)$ and a constant C_α, we consider the test

$$\Delta_{k,\alpha} := \mathbb{1}_{\left\{\hat{\mathrm{q}}^2_k \geq C_\alpha \frac{\nu^2_k}{n}\right\}} \qquad \text{with} \qquad \nu^2_k := \left(\sum_{j\in[\![k]\!]} \frac{1}{|\varphi_j|^4} \right)^{1/2}. \tag{3.1.7}$$

3.1.4 Minimax results

We show that for fixed $k \in \mathbb{N}$ the minimax risk of estimator $\hat{\mathrm{q}}^2_k$ defined in (3.1.6) is up to a constant bounded by

$$\rho^4_k \vee r^4_\circ \qquad \text{with} \qquad \rho^4_k := \left\{ a^4_k \vee \frac{\nu^4_k}{n^2} \right\} \qquad \text{and} \qquad r^4_\circ := \max_{m\in\mathbb{N}} \left\{ a^4_m \wedge \frac{a^2_m}{n\,|\varphi_m|^2} \right\}. \tag{3.1.8}$$

The base level term $r^4_\circ$ is present for all dimensions $k \in \mathbb{N}$, whereas the term ρ^4_k, which represents a typical bias^2-variance trade-off, explicitly depends on the dimension parameter $k \in \mathbb{N}$ and can, thus, be optimised with respect to k. More precisely, choosing $\kappa_\star$ as a minimizer of ρ^4_k, the risk of $\hat{\mathrm{q}}^2_{\kappa_\star}$ is up to a constant bounded by

$$\rho^4_\star \vee r^4_\circ := \left\{ \min_{k\in\mathbb{N}} \rho^4_k \right\} \vee r^4_\circ. \tag{3.1.9}$$

The term $r_\circ^4$ causes the classical elbow effect in quadratic functional estimation, since it prevents the rate from being faster than parametric. The upper bound shows the expected behaviour: a faster decay of the Fourier coefficients of φ, i.e. a smoother error density, results in a slower rate. Therefore, we call the decay of $(|\varphi_j|)_{j\in\mathbb{N}}$ the *degree of ill-posedness* of the model. On the other hand, a faster decay of the Fourier coefficients of the density of interest f, characterized by the sequence $a_\bullet$, yields a faster rate. We use the estimation upper bound to determine an upper bound for a radius of testing of the test $\Delta_{k,\alpha}$ defined in (3.1.7). For appropriately chosen C_α an upper bound for the radius of testing of $\Delta_{k,\alpha}$ is given by

$$\rho_k^2 = a_k^2 \vee \frac{\nu_k^2}{n}, \tag{3.1.10}$$

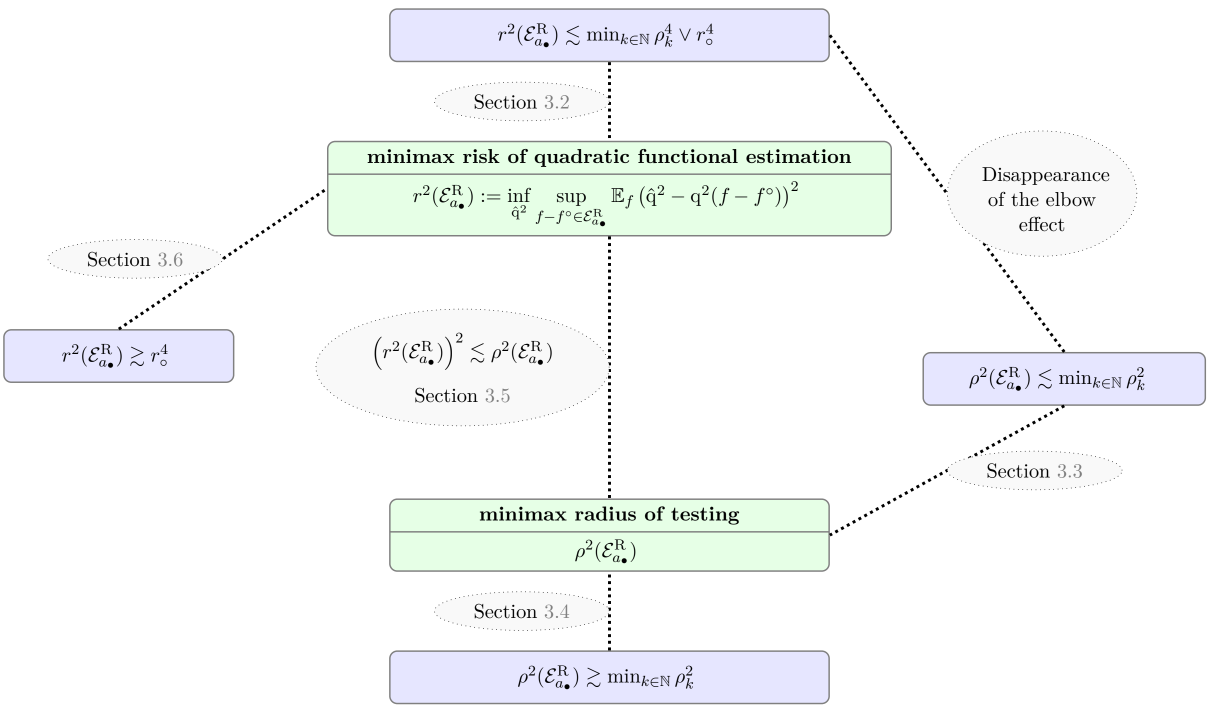
which can again be optimised with respect to $k \in \mathbb{N}$. Again choosing $\kappa_\star$ as the minimiser of ρ_k^2 with respect to k, the radius of testing of $\Delta_{\kappa_\star,\alpha}$ is of order

$$\rho_\star^2 := \min_{k\in\mathbb{N}} \rho_k^2.$$

Interestingly, the term causing the elbow effect in the estimation case vanishes in the radius of testing. Roughly speaking, the densities that cause $r_\circ^4$ in (3.1.9), and, hence, the elbow effect, are difficult to estimate (since they have large energy), but easy to test (since they are far from the null). This observation is explicitly used in the proof of the upper bounds of testing.

Outline

Minimax testing and quadratic functional estimation for circular convolution



Outline of this chapter. We provide an upper bound for the estimation risk in Section 3.2, which is used to derive an upper bound for the radius of testing in Section 3.3. Interestingly, the term causing the elbow effect disappears. Section 3.4 shows a matching lower bound for the testing problem. In Section 3.5 we show that testing is faster than quadratic functional estimation if we correct for the missing square, formally $r^4(\mathcal{E}) \geq C\rho^2(\mathcal{E})$ for some $C > 0$. Using this connection between quadratic functional estimation and testing, we immediately obtain a lower bound for the estimation problem. It remains to prove an additional lower bound for the term $r_\circ^4$ that causes the elbow effect, which is done in Section 3.6. Thus, we establish the order of both the minimax estimation risk and the minimax radius of testing. Technical results and their proofs are deferred to Appendix C.

3.2 Upper bound for the estimation risk

The next proposition presents an upper bound for the quadratic functional estimator defined in (3.1.6) for arbitrary $f \in \mathcal{D}$ and $k \in \mathbb{N}$. The key element of the proof is rewriting the estimator as a U-statistic and exploiting a well-known formula for its variance. The upper bound consists of one bias^2 term and two variance terms, one of which still involves the density of interest f.

Proposition 3.2.1 (Upper bound for the estimation risk). For $n \geq 2$ and $k \in \mathbb{N}$ the estimator defined in (3.1.6) satisfies

$$\mathbb{E}_f \left(\hat{\mathrm{q}}_k^2 - \mathrm{q}^2(f - f^\circ) \right)^2 \leq \left(\sum_{|j|>k} |f_j|^2 \right)^2 + \frac{c}{n^2} \sum_{|j| \in [\![k]\!]} \frac{1}{|\varphi_j|^4} + \frac{c}{n} \sum_{|j| \in [\![k]\!]} \frac{|f_j|^2}{|\varphi_j|^2} \tag{3.2.1}$$

with $c := \|f \circledast \varphi\|_\infty = \|g\|_\infty := \sup_{x \in [0,1)} |g(x)|$.

Proof of Proposition 3.2.1 . The bound follows from a classical bias^2-variance decomposition of the estimation risk;

$$\begin{aligned}
\mathbb{E}_f \left(\hat{\mathrm{q}}_k^2 - \mathrm{q}^2(f - f^\circ) \right)^2 &= \left(\mathbb{E}_f \hat{\mathrm{q}}_k^2 - \mathrm{q}^2(f - f^\circ) \right)^2 + \mathrm{var}_f(\hat{\mathrm{q}}_k^2) \\
&= \left(\mathrm{q}_k^2(f - f^\circ) - \mathrm{q}^2(f - f^\circ) \right)^2 + \mathrm{var}_f(\hat{\mathrm{q}}_k^2) \\
&= \left(\sum_{|j|>k} |f_j|^2 \right)^2 + \mathrm{var}_f(\hat{\mathrm{q}}_k^2).
\end{aligned} \tag{3.2.2}$$

To bound the variance, we rewrite the estimator as a U-statistic

$$\hat{\mathrm{q}}_k^2 = \frac{1}{n(n-1)} \sum_{l \neq m} \sum_{|j| \in [\![k]\!]} \frac{e_j(-Y_l) e_j(Y_m)}{|\varphi_j|^2} := \frac{1}{n(n-1)} \sum_{l \neq m} h(Y_l, Y_m) := \frac{1}{2} \mathrm{U}_n,$$

where we define the kernel $h : [0,1) \times [0,1) \longrightarrow \mathbb{C}$ by

$$h(y_1, y_2) := \sum_{|j| \in [\![k]\!]} \frac{e_j(-y_1) e_j(y_2)}{|\varphi_j|^2} \qquad \text{for } y_1, y_2 \in [0,1)$$

and the normalized U-statistic

$$\mathrm{U}_n := \binom{n}{2}^{-1} \sum_{l \neq m} h(Y_l, Y_m).$$

The kernel h is symmetric and real-valued. Indeed, for $y_1, y_2 \in [0,1)$ we have

$$h(y_1, y_2) = \sum_{|j| \in [\![k]\!]} \frac{e_j(-y_1) e_j(y_2)}{|\varphi_j|^2} = \sum_{|l| \in [\![k]\!]} \frac{e_l(y_1) e_l(-y_2)}{|\varphi_j|^2} = h(y_2, y_1),$$

where we introduce the change of variables $l=-j$ and exploit that $e_j(\cdot)=e_{-j}(-\cdot)$. Moreover,

$$\overline{h(y_1,y_2)}=\sum_{|j|\in[\![k]\!]}\frac{e_j(y_1)e_j(-y_2)}{|\varphi_j|^2}=\sum_{|l|\in[\![k]\!]}\frac{e_l(-y_1)e_l(y_2)}{|\varphi_j|^2}=h(y_1,y_2),$$

where we again introduce the change of variables $l=-j$ and exploit that $\overline{e_j(\cdot)}=e_j(-\cdot)$. Let us define the function

$$h_1:[0,1)\longrightarrow\mathbb{R},y\longmapsto h_1(y):=\mathbb{E}_f(h(y,Y_2)).$$

By Lemma A on p. 183 in Serfling [2009] the variance of the U-statistic U_n is determined by

$$\mathrm{var}_f(\mathrm{U}_n)=\binom{n}{2}^{-1}(2(n-2)\xi_1+\xi_2)$$

with $\xi_1:=\mathrm{var}_f(h_1(Y_1))$ and $\xi_2:=\mathrm{var}_f(h(Y_1,Y_2))$. Next, we bound the two terms ξ_1 and ξ_2. Since

$$h_1(y)=\mathbb{E}_f(h(y,Y_2))=\sum_{|j|\in[\![k]\!]}\frac{\mathbb{E}_f e_j(Y_2)}{|\varphi_j|^2}e_j(-y)=\sum_{|j|\in[\![k]\!]}\frac{\overline{g_j}}{|\varphi_j|^2}e_j(-y),$$

we obtain by Parseval's identity

$$\xi_1\leq\mathbb{E}_f\,|h_1(Y_1)|^2\leq\|g\|_\infty\,\|h_1\|^2_{\mathscr{L}^2}=\|g\|_\infty\sum_{|j|\in[\![k]\!]}\frac{|f_j|^2}{|\varphi_j|^2}.$$

Now consider the term ξ_2. It holds

$$\xi_2=\mathrm{var}_f(h(Y_1,Y_2))\leq\mathbb{E}_f\,|h(Y_1,Y_2)|^2\leq\|g\|_\infty\int_{[0,1)}\int_{[0,1)}|h(y_1,y_2)|^2\,\mathrm{d}y_1 g(y_2)\mathrm{d}y_2,$$

where

$$\int|h(y_1,y_2)|^2\,\mathrm{d}y_1=\sum_{|j|,|l|\in[\![k]\!]}\frac{\int_{[0,1)}e_j(y_2-y_1)\overline{e_l(y_2-y_1)}\mathrm{d}y_1}{|\varphi_j|^2\,|\varphi_l|^2}=\sum_{|j|\in[\![k]\!]}\frac{1}{|\varphi_j|^4}$$

and, hence,

$$\int_{[0,1)}\int_{[0,1)}|h(y_1,y_2)|^2\,\mathrm{d}y_1 g(y_2)\mathrm{d}y_2=\sum_{|j|\in[\![k]\!]}\frac{1}{|\varphi_j|^4}\int_{[0,1)}g(y_2)\mathrm{d}y_2=\sum_{|j|\in[\![k]\!]}\frac{1}{|\varphi_j|^4}.$$

Finally, combining the bounds for ξ_1 and ξ_2 yields

$$\begin{aligned}\mathrm{var}_f(\hat{\mathrm{q}}^2_k)=\frac{1}{4}\,\mathrm{var}_f(\mathrm{U}_n)&=\frac{2(n-2)\xi_1+\xi_2}{2n(n-1)}\leq\frac{1}{n}\xi_1+\frac{1}{n^2}\xi_2\\&\leq\frac{\|g\|_\infty}{n}\sum_{|j|\in[\![k]\!]}\frac{|f_j|^2}{|\varphi_j|^2}+\frac{\|g\|_\infty}{n^2}\sum_{|j|\in[\![k]\!]}\frac{1}{|\varphi_j|^4},\end{aligned}\tag{3.2.3}$$

where we use that $\frac{1}{2(n-1)}\leq\frac{1}{n}$ for $n\geq 2$. Together with (3.2.2) this proves the assertion. □

The upper bound in (3.2.1) depends on the quantity $c=\|g\|_\infty\leq\|\varphi\|_\infty$, which is uniformly bounded for all $f\in\mathcal{D}$ as soon as $\|\varphi\|_\infty<\infty$. By additionally exploiting the regularity condition (3.1.4) we obtain a uniform bound for the risk, valid for all $f\in\mathcal{E}^{\mathrm{R}}_{a_\bullet}$.

Corollary 3.2.2 (Uniform upper bound for the risk of estimation).
Consider the quantities ν_k^4 and $r_\circ^4$ as defined in (3.1.7) and (3.1.8), respectively. For $n, k \in \mathbb{N}$, $n \geq 2$ the estimator $\hat{\mathrm{q}}_k^2$ defined in (3.1.6) satisfies

$$\sup_{f \in \mathcal{E}_{a_\bullet}^{\mathrm{R}}} \mathbb{E}_f(\hat{\mathrm{q}}_k^2 - \mathrm{q}^2(f - f^\circ))^2 \leq \quad c_1 a_k^4 \quad \vee \quad c_2 \frac{\nu_k^4}{n^2} \quad \vee \quad c_3 r_\circ^4 \tag{3.2.4}$$

with $c_1 := 3\mathrm{R}^4$, $c_2 := 2(\|\varphi\|_\infty + \mathrm{R}^2)$, $c_3 := 3\,\|\varphi\|_\infty\,\mathrm{R}^2$.

Proof of Corollary 3.2.2. We exploit the upper bound of Proposition 3.2.1. Since the sequence $a_\bullet$ is non-increasing, the first term on the right-hand side in (3.2.1) (the bias term) is bounded by

$$\sum_{|j|>k} |f_j|^2 = 2\sum_{j>k} |f_j|^2 = 2\sum_{j>k} |f_j|^2\, a_j^{-2} a_j^2 \leq 2a_k^2 \sum_{j>k} |f_j|^2\, a_j^{-2} \leq \mathrm{R}^2 a_k^2.$$

To control the last (third) term on the right-hand side of (3.2.1), we bound each summand, i.e. for each $j \in \mathbb{N}$ we have

$$\frac{1}{n}\frac{|f_j|^2}{|\varphi_j|^2} \leq \begin{cases} \frac{|f_j|^2}{a_j^2} a_j^4 \left(1 \wedge \frac{1}{n|\varphi_j|^2 a_j^2}\right) & \text{if } n\,|\varphi_j|^2\, a_j^2 \geq 1, \\ \frac{\mathrm{R}^2}{n^2 |\varphi_j|^4} & \text{if } n\,|\varphi_j|^2\, a_j^2 < 1. \end{cases}$$

Hence, we obtain a bound for the entire sum

$$\begin{aligned}\frac{1}{n}\sum_{|j|\in[\![k]\!]} \frac{|f_j|^2}{|\varphi_j|^2} &\leq \sum_{|j|\in[\![k]\!]} \frac{|\varphi_j|^2}{a_j^2} a_j^4 \left(1 \wedge \frac{1}{n\,|\varphi_j|^2\, a_j^2}\right) + \frac{\mathrm{R}^2}{n^2} \sum_{|j|\in[\![k]\!]} \frac{1}{|\varphi_j|^4} \\ &\leq \max_{m\in\mathbb{N}} \left\{ a_m^4 \wedge \frac{a_m^2}{n\,|\varphi_m|^2} \right\} \sum_{|j|\in[\![k]\!]} \frac{|f_j|^2}{a_j^2} + \mathrm{R}^2 \frac{\nu_k^4}{n^2} \\ &\leq \mathrm{R}^2 r_\circ^4 + \mathrm{R}^2 \frac{\nu_k^4}{n^2}.\end{aligned}$$

Combining both bounds with (3.2.1) and $x + y + z \leq 3(x \vee y \vee z)$ for all $x, y, z \geq 0$ yields the assertion. □

We remind the reader that since we consider the case $f^\circ = \mathbb{1}_{[0,1)}$, we have $f \in \mathcal{E}_{a_\bullet}^{\mathrm{R}}$ if and only if $f - f^\circ \in \mathcal{E}_{a_\bullet}^{\mathrm{R}}$.

Remark 3.2.3 (Optimal choice of the dimension parameter). *The first two terms in the upper bound of Corollary 3.2.2 depend on the dimension parameter $k \in \mathbb{N}$, whereas the last term $c_3 r_\circ^4$ does not. It plays the role of a base-level error, which causes the well-known elbow effect in quadratic functional estimation (cp. also Illustration 3.2.6 below). It can easily be seen that $\mathrm{r}_\circ^4$ is always of order larger than $\frac{1}{n}$, since $\mathrm{r}_\circ^4 = \max_{m\in\mathbb{N}} \left\{ a_m^4 \wedge \frac{a_m^2}{n|\varphi_m|^2} \right\} \geq a_1^4 \wedge \frac{a_1^2}{n|\varphi_j|^2} \gtrsim \frac{1}{n}$. In other words, no matter the choice of the dimension k the estimation rate can never be faster than parametric. The first two terms, however, depend on $k \in \mathbb{N}$ and can, therefore, be optimised. We define the optimal dimension*

$$\kappa_\star := \min\left\{ k \in \mathbb{N} : a_k^4 \leq \frac{\nu_k^4}{n^2} \right\} \tag{3.2.5}$$

as the dimension that achieves an optimal bias-variance trade-off. □

Theorem 3.2.4 (Upper bound for the minimax risk of estimation). Let $\kappa_\star$ as in (3.2.5), $\rho_\star$ as in (3.1.9) and $r_\circ^4$ as in (3.1.8). For $n \geq 2$ the minimax risk satisfies

$$r^2(\mathcal{E}_{a_\bullet}^{\mathrm{R}}) \leq r^2(\hat{\mathrm{q}}_{\kappa_\star}^2, \mathcal{E}_{a_\bullet}^{\mathrm{R}}) \leq C \left(\rho_\star^4 \vee r_\circ^4\right)$$

with $C := 3\,(\mathrm{R}^4 + \|\varphi\|_\infty + \mathrm{R}^2 + \|\varphi\|_\infty \mathrm{R}^2)$.

Proof of Theorem 3.2.4. We apply Corollary 3.2.2 to $\hat{\mathrm{q}}_{\kappa_\star}^2$ with $\kappa_\star$ as in (3.2.5). □

We now provide an additional upper bound for the variance of the estimator (3.1.6), which is used in the next section to derive an upper bound for the testing radius.

Corollary 3.2.5 (Upper bound for the variance). Let $f^\circ = \mathbb{1}_{[0,1)}$ and $f \in \mathcal{D}$. For $n, k \in \mathbb{N}$, $n \geq 2$ and ν_k^2 as in (3.1.7) the estimator defined in (3.1.6) satisfies

$$\mathrm{var}_{f^\circ}(\hat{\mathrm{q}}_k^2) \leq \frac{\nu_k^4}{n^2}, \tag{3.2.6}$$

$$\mathrm{var}_f(\hat{\mathrm{q}}_k^2) \leq \|\varphi\|_\infty \cdot \mathrm{q}_k^2(f - f^\circ)\frac{\nu_k^2}{n} + \|\varphi\|_\infty \frac{\nu_k^4}{n^2}. \tag{3.2.7}$$

Proof of Corollary 3.2.5. Let us start with the second assertion (3.2.7). We use the bound (3.2.3) derived in the proof of Proposition 3.2.1 combined with $\|g\|_\infty \leq \|\varphi\|_\infty$. The first term on the right hand side can be bounded due to the Cauchy-Schwarz inequality by

$$\sum_{|j| \in [\![k]\!]} \frac{|f_j|^2}{|\varphi_j|^2} \leq \left(\sum_{|j| \in [\![k]\!]} |f_j|^4\right)^{1/2} \left(\sum_{|j| \in [\![k]\!]} \frac{1}{|\varphi_j|^4}\right)^{1/2} \leq \mathrm{q}_k^2(f - f^\circ) \cdot \nu_k^2.$$

In the last inequality we exploit $\sqrt{x+y} \leq \sqrt{x} + \sqrt{y}$ for any $x, y \geq 0$, which shows (3.2.7). To prove the first assertion (3.2.6), we note that, additionally, for $f = f^\circ = \mathbb{1}_{[0,1)}$ and, hence, $g = \mathbb{1}_{[0,1)}$ we have $\|g\|_\infty = 1$ and $\mathrm{q}_k^2(f - f^\circ) = 0$, which finishes the proof. □

Illustration 3.2.6. We illustrate the order of the estimation risk under typical regularity and ill-posedness assumptions. For two real-valued sequences $(x_j)_{j\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ and $(y_j)_{j\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ we write $x_j \lesssim y_j$ if there exists a constant $c > 0$ such that $x_j \leq c y_j$ for all $j \in \mathbb{N}$. We write $x_j \sim y_j$, if both $x_j \lesssim y_j$ and $y_j \lesssim x_j$. Concerning the class $\mathcal{E}_{a_\bullet}^{\mathrm{R}}$ we distinguish two behaviours of the sequence $a_\bullet$, namely the **ordinary smooth** case, $a_j \sim j^{-s}$ for $s > 1/2$, corresponding to a Sobolev ellipsoid, and the **super smooth** case, $a_j \sim \exp(-j^s)$ for $s > 0$, corresponding to a class of analytic functions. We also distinguish two cases for the regularity of the error density φ. For $p > 1/2$ we consider a **mildly ill-posed** model $|\varphi_j| \sim |j|^{-p}$ and for $p > 0$ a **severely ill-posed model** $|\varphi_j| \sim \exp(-|j|^p)$. Many examples of circular densities can be found in Chapter 3 of Mardia and Jupp [2009]. The table below presents the order of the upper bound for the minimax risk $r^2(\mathcal{E}_{a_\bullet}^{\mathrm{R}})$ derived in Theorem 3.2.4. In Section 3.4 we provide a matching lower bound, and, thus, establish the minimax optimality of the estimator $\hat{\mathrm{q}}_{\kappa_\star}^2$. Note that in the mildly ill-posed – ordinary smooth case we observe the typical elbow effect in quadratic functional estimation.

Order of the minimax risk of estimation $r^2(\mathcal{E}^{\mathrm{R}}_{a_\bullet}) \sim \rho_\star^4 \vee r_\circ^4$				
a_j (smooth.)	$\lvert\varphi_j\rvert$ (ill-posed.)	$\rho_\star^4$	$r_\circ^4$	$r^2(\mathcal{E}^{\mathrm{R}}_{a_\bullet})$
j^{-s}	$\lvert j\rvert^{-p}$	$n^{-\frac{8s}{4s+4p+1}}$	$\begin{cases} n^{-\frac{8s}{4s+4p}} & s-p<0 \\ n^{-1} & s-p\geq 0 \end{cases}$	$\begin{cases} n^{-\frac{8s}{4s+4p+1}} & s-p<\frac14 \\ n^{-1} & s-p\geq \frac14 \end{cases}$
j^{-s}	$e^{-\lvert j\rvert^p}$	$(\log n)^{-\frac{4s}{p}}$	$(\log n)^{-\frac{4s}{p}}$	$(\log n)^{-\frac{4s}{p}}$
e^{-j^s}	$\lvert j\rvert^{-p}$	$n^{-2}(\log n)^{\frac{4p+1}{s}}$	n^{-1}	n^{-1}

Calculations for the risk bounds in Illustration 3.2.6.
We establish the order of the terms $r_\circ^4$ and $\rho_\star^4$ in Theorem 3.2.4 for each of the three combinations in Illustration 3.2.6 and determine the dominating term. Let $m_\star := \max\left\{m\in\mathbb{N} : a_m^4 \geq \frac{a_m^2}{n|\varphi_m|^2}\right\}$.

1. **(ordinary smooth – mildly ill-posed)** Consider first $\rho_\star^4$ defined in (3.1.9). The variance term $\frac{\nu_k^4}{n^2} = \frac{1}{n^2}\sum_{|j|\in[\![k]\!]}\frac{1}{|\varphi_j|^4} \sim \frac{1}{n^2}\sum_{|j|\in[\![k]\!]}|j|^{4p}$ is of order $\frac{1}{n^2}k^{4p+1}$ and the bias term a_k^4 is of order k^{-4s}. Hence, the optimal $\kappa_\star$ satisfies $\kappa_\star^{-4s}\sim\frac{1}{n^2}\kappa_\star^{4p+1}$ and, thus, $\kappa_\star\sim n^{\frac{2}{4s+4p+1}}$, which yields an upper bound of order $\rho_\star^4\sim\kappa_\star^{-4s}\sim n^{-\frac{8s}{4s+4p+1}}$.
   For the base level $r_\circ^4=\max_{m\in\mathbb{N}}\left\{a_m^4\wedge\frac{a_m^2}{n|\varphi_j|^2}\right\}$, the term $\frac{a_m^2}{n|\varphi_m|^2}\sim\frac1n m^{2(p-s)}$ is monotonically increasing in m for $p-s>0$ and monotonically non-increasing otherwise. Let $p-s>0$, then $m_\star$ satisfies $m_\star^{-4s}\sim\frac1n m_\star^{2(p-s)}$ and is thus of order $m_\star\sim n^{\frac{2s}{s+p}}$. Therefore, $r_\circ^4\sim n^{-\frac{8s}{4s+4p}}$ is negligible compared with $\rho_\star^4$. Let $p-s\leq 0$, then a_m^4 and $\frac{a_m^2}{n|\varphi_m|^2}$ are non-increasing. The maximum of their minimum is attained at $m_\star=1$, which yields $r_\circ^4\sim\frac1n$. Hence, $r_\circ^4$ is of larger order than $\rho_\star^4$ for $s-p>\frac14$ only.

2. **(ordinary smooth – severely ill-posed)** Consider first $\rho_\star^4$ defined in (3.1.9). The variance term $\frac{\nu_k^4}{n^2}=\frac{1}{n^2}\sum_{|j|\in[\![k]\!]}\frac{1}{|\varphi_j|^4}\sim\frac{1}{n^2}\sum_{|j|\in[\![k]\!]}\exp(4\,|j|^p)$ is of order $\frac{1}{n^2}\exp(4k^p)$ and the bias term a_k^4 is of order k^{-4s}. Hence, the optimal $\kappa_\star$ satisfies $\kappa_\star^{-4s}\sim\frac{1}{n^2}\exp(4\kappa_\star^p)$ and, thus, $\kappa_\star\sim(\log(n^2/b_n))^{\frac1p}$ with $b_n\sim(\log(n^2))^{\frac{4s}{p}}$, which yields an upper bound of order $\rho_\star^4\sim\kappa_\star^{-4s}\sim(\log n)^{-\frac{4s}{p}}$.
   Considering the base level $r_\circ^4=\max_{m\in\mathbb{N}}\left\{a_m^4\wedge\frac{a_m^2}{n|\varphi_j|^2}\right\}$, the term $\frac{a_m^2}{n|\varphi_m|^2}\sim\frac{m^{-2s}}{n}\exp(2m^p)$ is eventually monotonically increasing in m. Hence, $m_\star$ satisfies $m_\star^{-4s}\sim\frac1n m_\star^{-2s}\exp(2m_\star^p)$ and is thus of order $m_\star\sim(\log(n/b_n))^{\frac1p}$ with $b_n\sim(\log n)^{\frac{2s}{p}}$. Therefore, $r_\circ^4\sim(\log n)^{-\frac{4s}{p}}$ is of the same order as $\rho_\star^4$.

3. **(super smooth – mildly ill-posed)** Consider first $\rho_\star^4$ defined in (3.1.9). The variance term $\frac{\nu_k^4}{n^2}=\frac{1}{n^2}\sum_{|j|\in[\![k]\!]}\frac{1}{|\varphi_j|^4}\sim\frac{1}{n^2}\sum_{|j|\in[\![k]\!]}|j|^{4p}$ is of order $\frac{1}{n^2}k^{4p+1}$ and the bias term a_k^4 is of order $\exp(-4k^s)$. Hence, the optimal $\kappa_\star$ satisfies $\exp(-4\kappa_\star^s)\sim\frac{1}{n^2}\kappa_\star^{4p+1}$ and, thus, $\kappa_\star\sim(\log(n^2/b_n))^{\frac1s}$ with $b_n\sim(\log n)^{\frac{4p+1}{s}}$, which yields an upper bound of order $\rho_\star^4\sim\frac{1}{n^2}\kappa_\star^{4p+1}\sim\frac{1}{n^2}(\log n)^{\frac{4p+1}{s}}$.
   Considering the base level $r_\circ^4=\max_{m\in\mathbb{N}}\left\{a_m^4\wedge\frac{a_m^2}{n|\varphi_j|^2}\right\}$, the term $\frac{a_m^2}{n|\varphi_m|^2}\sim\frac{m^{2p}}{n}\exp(-2m^s)$ is eventually monotonically decreasing in m. Hence, $m_\star$ satisfies $m_\star\sim 1$. Therefore, $r_\circ^4\sim n^{-1}$ is of larger order than $\rho_\star^4$ and is thus the dominant term.

□

3.3 Upper bound for the radius of testing

In this section we derive an upper bound for the radius of testing of the task (3.1.2). For $k \in \mathbb{N}$ we consider the family of tests $\{\Delta_{k,\alpha}\}_{\alpha\in(0,1)}$ defined in (3.1.7), that is based on the estimator $\hat{\mathrm{q}}_k^2$ in (3.1.6) of the distance $\|f - f^\circ\|^2_{\mathscr{L}^2}$ to the null hypothesis.

Proposition 3.3.1 (Upper bound for the radius of testing of $\Delta_{k,\alpha/2}$). Let $\alpha \in (0,1)$, $c := \|\varphi\|_\infty$ and $C_{\alpha/2}, \widetilde{A}_\alpha$ be such that

$$c \cdot \frac{2C_{\alpha/2}+1}{C^2_{\alpha/2}} \le \frac{\alpha}{2} \qquad \text{and} \qquad c \cdot \frac{2C_{\alpha/2}+1}{(\widetilde{A}_\alpha - C_{\alpha/2})^2} \le \frac{\alpha}{2}. \tag{3.3.1}$$

Set $\overline{A}^2_\alpha := \mathrm{R}^2 + \widetilde{A}^2_\alpha$. Then, for all $A \ge \overline{A}_\alpha$ and all $k \in \mathbb{N}$ we obtain

$$\mathcal{R}\left(\Delta_{k,\alpha/2} \mid \mathcal{E}^{\mathrm{R}}_{a_\bullet}, A\rho_k\right) \le \alpha,$$

i.e. $\rho_k^2 = a_k^2 \vee \frac{\nu_k^2}{n}$ is an upper bound for the radius of testing of $\left\{\Delta_{k,\alpha/2}\right\}_{\alpha\in(0,1)}$.

Remark 3.3.2 (Choice of $C_{\alpha/2}$ and $\widetilde{A}_\alpha$). *In particular* (3.3.1) *and, hence, Proposition 3.3.1 is satisfied for*

$$C_{\alpha/2} = 2(\alpha/2)^{-1}\|\varphi\|_\infty \qquad \textit{and} \qquad \widetilde{A}_\alpha = C_{\alpha/2} + \sqrt{2/\alpha}\sqrt{4\|\varphi\|^2_\infty(\alpha/2)^{-1} + \|\varphi\|_\infty}.$$

Indeed, since $c \ge 1$ we have

$$c \cdot \frac{2C_{\alpha/2}+1}{C^2_{\alpha/2}} = c \cdot \left(\frac{1}{2c}\frac{\alpha}{2} + \frac{1}{4c^2}\left(\frac{\alpha}{2}\right)^2\right) \le \frac{\alpha}{2}\left(\frac{1}{2} + \frac{1}{4c}\right) \le \frac{\alpha}{2}$$

and

$$c \cdot \frac{2C_{\alpha/2}+1}{(\widetilde{A}_\alpha - C_{\alpha/2})^2} = c \cdot \frac{\alpha}{2} \cdot \frac{2C_{\alpha/2}+1}{4c^2(\alpha/2)^{-1} + c} = \frac{\alpha}{2} \cdot \frac{2C_{\alpha/2}+1}{4c(\alpha/2)^{-1}+1} = \frac{\alpha}{2}.$$

□

Proof of Proposition 3.3.1. We show that both the type I error probability and the type II error probability are bounded by $\alpha/2$. Consider first the **type I error probability**. Applying first Markov's inequality and then the second inequality (3.2.6) from Corollary 3.2.5 we obtain

$$\begin{aligned}\mathbb{P}_{f^\circ}(\Delta_{k,\alpha/2} = 0) &= \mathbb{P}_{f^\circ}(\hat{\mathrm{q}}_k^2 \ge C_{\alpha/2}\frac{\nu_k^2}{n}) \\ &\le \frac{\mathbb{E}_{f^\circ}\left(\hat{\mathrm{q}}_k^2\right)^2}{C^2_{\alpha/2}n^{-2}\nu_k^4} = \frac{\mathrm{var}_{f^\circ}(\hat{\mathrm{q}}_k^2)}{C^2_{\alpha/2}n^{-2}\nu_k^4} \le \frac{1}{C^2_{\alpha/2}} \le \frac{\alpha}{2},\end{aligned} \tag{3.3.2}$$

for all $C_{\alpha/2}$ satisfying (3.3.1), since $\|\varphi\|_\infty \ge 1$. Next, we consider the **type II error probability**. Let f be contained in the $\overline{A}_\alpha\rho_k$-separated alternative, i.e. $f - f^\circ \in \mathcal{E}^{\mathrm{R}}_{a_\bullet}$ and $\mathrm{q}^2(f - f^\circ) \ge (\overline{A}_\alpha)^2\rho_k^2$. We expand

$$\mathbb{P}_f(\Delta_{k,\alpha/2} = 0) = \mathbb{P}_f(\hat{\mathrm{q}}_k^2 < C_{\alpha/2}\frac{\nu_k^2}{n}) = \mathbb{P}_f(\hat{\mathrm{q}}_k^2 - \mathrm{q}_k^2(f - f^\circ) < C_{\alpha/2}\nu_k^2 - \mathrm{q}_k^2(f - f^\circ))$$

and distinguish the following two cases for the density f

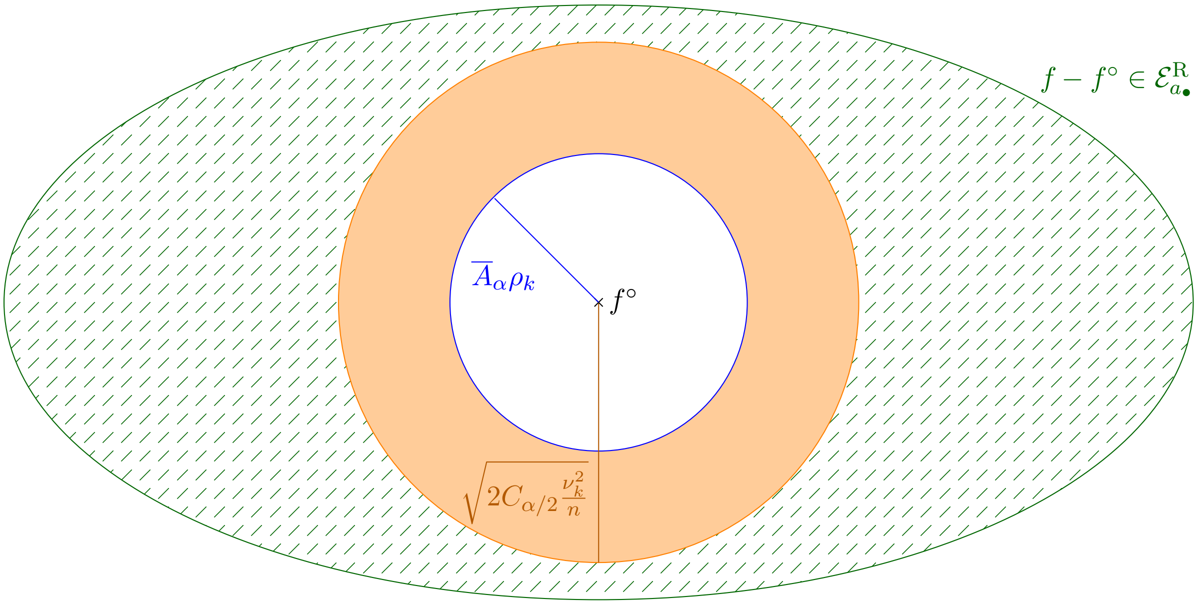


Figure 3.1: **Visualization of the structure of the proof of Proposition 3.3.1.** We distinguish the two cases: Either $f - f^\circ$ has large energy (in the first k components), hence, it is easy to test since it is far from the null. Or $f - f^\circ$ has small energy (in the first k components), hence, it is difficult to test since it is close to the null.

1. $\mathrm{q}_k^2(f - f^\circ) \geq 2C_{\alpha/2}\frac{\nu_k^2}{n}$, (easy to test)
2. $\mathrm{q}_k^2(f - f^\circ) < 2C_{\alpha/2}\frac{\nu_k^2}{n}$. (difficult to test)

Case 1: (easy to test) We have $C_{\alpha/2}\frac{\nu_k^2}{n} - \mathrm{q}_k^2(f - f^\circ) \leq -\frac{1}{2}\mathrm{q}_k^2(f - f^\circ)$ and, therefore, due to Markov's inequality

$$\begin{aligned}\mathbb{P}_f(\Delta_{k,\alpha/2} = 0) &\leq \mathbb{P}_f(\hat{\mathrm{q}}_k^2 - \mathrm{q}_k^2(f - f^\circ) \leq -\frac{1}{2}\mathrm{q}_k^2(f - f^\circ)) \\ &= \mathbb{P}_f(\mathrm{q}_k^2(f - f^\circ) - \hat{\mathrm{q}}_k^2 \geq \frac{1}{2}\mathrm{q}_k^2(f - f^\circ)) \leq 4\frac{\mathrm{var}_f(\hat{\mathrm{q}}^2)}{(\mathrm{q}_k^2(f - f^\circ))^2}.\end{aligned}$$

On the one hand, by the case distinction we have $\mathrm{q}_k^2(f - f^\circ) \geq 2C_{\alpha/2}\frac{\nu_k^2}{n}$, on the other hand we have $\mathrm{var}_f(\hat{\mathrm{q}}^2) \leq c\mathrm{q}_k^2(f - f^\circ)\frac{\nu_k^2}{n} + c\frac{\nu_k^4}{n^2}$ with $c = \|\varphi\|_\infty$ due to (3.2.7) in Corollary 3.2.5. Hence,

$$\begin{aligned}\mathbb{P}_f(\Delta_{k,\alpha/2} = 0) &\leq 4\frac{c\mathrm{q}_k^2(f - f^\circ)\frac{\nu_k^2}{n} + c\frac{\nu_k^4}{n^2}}{(\mathrm{q}_k^2(f - f^\circ))^2} = 4c\left(\frac{\frac{\nu_k^2}{n}}{\mathrm{q}_k^2(f - f^\circ)} + \frac{\frac{\nu_k^4}{n^2}}{(\mathrm{q}_k^2(f - f^\circ))^2}\right) \\ &\leq 4c\left(\frac{\frac{\nu_k^2}{n}}{2C_{\alpha/2}\frac{\nu_k^2}{n}} + \frac{\frac{\nu_k^4}{n^2}}{4C_{\alpha/2}^2\frac{\nu_k^4}{n^2}}\right) = c \cdot \frac{2C_{\alpha/2} + 1}{C_{\alpha/2}^2} \leq \frac{\alpha}{2}\end{aligned}$$

due to assumption (3.3.1).

Case 2: (difficult to test) Under the alternative we have

$$\sum_{|j|>k} |f_j|^2 = 2\sum_{j>k} |f_j|^2 = 2\sum_{j>k} |f_j|^2 a_j^{-2} a_j^2 \leq 2a_k^2 \sum_{j>k} |f_j|^2 a_j^{-2} \leq \mathrm{R}^2 a_k^2.$$

and $\mathrm{q}^2(f - f^\circ) = \sum_{|j|\in\mathbb{N}} |f_j|^2 \geq \overline{A}_\alpha^2 \rho_k^2$. Therefore, it follows

$$\mathrm{q}_k^2(f - f^\circ) = \mathrm{q}^2(f - f^\circ) - \sum_{|j|>k} |f_j|^2 \geq \overline{A}_\alpha^2 \rho_k^2 - a_k^2 \mathrm{R}^2 = \widetilde{A}_\alpha^2 \frac{\nu_k^2}{n} + \mathrm{R}^2 a_k^2 - \mathrm{R}^2 a_k^2 = \widetilde{A}_\alpha^2 \frac{\nu_k^2}{n}.$$

Hence, due to Markov's inequality the type II error probability satisfies

$$\begin{aligned}
\mathbb{P}_f(\Delta_{k,\alpha/2}=0) &= \mathbb{P}_f\left(\hat{\mathrm{q}}_k^2 - \mathrm{q}_k^2(f-f^\circ) \leq C_{\alpha/2}\frac{\nu_k^2}{n} - \mathrm{q}_k^2(f-f^\circ)\right)\\
&\leq \mathbb{P}_f\left(\hat{\mathrm{q}}_k^2 - \mathrm{q}_k^2(f-f^\circ) \leq \left(C_{\alpha/2} - \widetilde{A}_\alpha^2\right)\frac{\nu_k^2}{n}\right)\\
&= \mathbb{P}_f\left(-\hat{\mathrm{q}}_k^2 + \mathrm{q}_k^2(f-f^\circ) \leq \left(-C_{\alpha/2} + \widetilde{A}_\alpha^2\right)\frac{\nu_k^2}{n}\right)\\
&\leq \frac{\mathrm{var}_f(\hat{\mathrm{q}}_k^2)}{\left(\widetilde{A}_\alpha^2 - C_{\alpha/2}\right)^2\frac{\nu_k^4}{n^2}}.
\end{aligned}$$

From the bound for the variance (3.2.7), the case distinction condition and the choice of $\widetilde{A}_\alpha$ in (3.3.1) it follows

$$\begin{aligned}
\mathbb{P}_f(\Delta_{k,\alpha/2}=0) &\leq c\cdot\frac{\mathrm{q}_k^2(f-f^\circ)\frac{\nu_k^2}{n} + \frac{\nu_k^4}{n^2}}{\left(\widetilde{A}_\alpha^2 - C_{\alpha/2}\right)^2\frac{\nu_k^4}{n^2}} \leq c\cdot\frac{2C_{\alpha/2}\frac{\nu_k^2}{n} + \frac{\nu_k^2}{n}}{\left(\widetilde{A}_\alpha^2 - C_{\alpha/2}\right)^2\frac{\nu_k^2}{n}}\\
&= c\cdot\frac{2C_{\alpha/2}+1}{\left(\widetilde{A}_\alpha^2 - C_{\alpha/2}\right)^2} \leq \frac{\alpha}{2}.
\end{aligned}$$

Combining the last bound and (3.3.2), we obtain the assertion, which completes the proof. □

From Proposition 3.3.1 with $k=\kappa_\star$ as in (3.2.5) and $\rho_\star$ as in (3.1.9) we immediately obtain the following corollary and, hence, omit its proof.

Corollary 3.3.3 (Upper bound for the minimax radius of testing). Under the conditions of Proposition 3.3.1 for all $A \geq \overline{A}_\alpha$ we obtain

$$\mathcal{R}\left(\mathcal{E}_{a_\bullet}^{\mathrm{R}}, A\rho_\star\right) \leq \mathcal{R}\left(\Delta_{\kappa_\star,\alpha/2} \mid \mathcal{E}_{a_\bullet}^{\mathrm{R}}, A\rho_\star\right) \leq \alpha,$$

i.e. $\rho_\star^2 = \min_{k\in\mathbb{N}}\left\{a_k^2 \vee \frac{\nu_k^2}{n}\right\}$ is an upper bound for the minimax radius of testing.

Illustration 3.3.4. We illustrate the order of the upper bound for the radius of testing $\rho_\star^2 = \rho_\star^2(\mathcal{E}_{a_\bullet}^{\mathrm{R}})$ derived in Corollary 3.3.3 under the typical smoothness and ill-posedness assumptions introduced in Illustration 3.2.6. In the next section we provide a matching lower bound, which establishes $\rho_\star^2$ as the minimax radius of testing. Comparing the next table with Illustration 3.2.6 we emphasise that there is no elbow effect. The derivation of the bounds is similar to the ones in Illustration 3.2.6 and is thus omitted.

Order of the minimax radius of testing $\rho_\star^2(\mathcal{E}^{\mathrm{R}}_{a_\bullet})$		
a_j (smoothness)	$\lvert\varphi_j\rvert$ (ill-posedness)	$\rho_\star^2$
j^{-s}	$\lvert j\rvert^{-p}$	$n^{-\frac{4s}{4s+4p+1}}$
j^{-s}	$e^{-\lvert j\rvert^p}$	$(\log n)^{-\frac{2s}{p}}$
e^{-j^s}	$\lvert j\rvert^{-p}$	$n^{-1}(\log n)^{\frac{4p+1}{2s}}$

3.4 Lower bound for the radius of testing

In this section we prove a matching lower bound for the radius of testing. The proof is inspired by Assouad's cube technique (see Tsybakov [2009], Chapter 2.7 for an explanation of the technique in the estimation case), where the testing risk is reduced to a distance between probability measures. It requires the construction of $2^{\kappa_\star}$ candidates (called hypotheses) in the class $\mathcal{E}^{\mathrm{R}}_{a_\bullet}$, which are vertices of a hypercube. Roughly speaking, they are constructed such that they are statistically indistinguishable from the null f° while having largest possible $\mathscr{L}^2$-distance.

Proposition 3.4.1 (Lower bound for the radius of testing). Assume

$$2\sum_{j\in\mathbb{N}} a_j^2 =: \mathfrak{a} < \infty. \tag{3.4.1}$$

Consider $\kappa_\star$ as in (3.2.5) and let $\eta \in (0,1]$ satisfy

$$\left(a_{\kappa_\star}^2 \vee \frac{\nu_{\kappa_\star}^2}{n}\right)\eta \le \left(a_{\kappa_\star}^2 \wedge \frac{\nu_{\kappa_\star}^2}{n}\right). \tag{3.4.2}$$

For $\alpha \in (0,1)$ define $\underline{A}_\alpha^2 := \eta\left(\mathrm{R}^2 \wedge \sqrt{\log(1+2\alpha^2)} \wedge \mathfrak{a}^{-1}\right)$. Then, for all $A \le \underline{A}_\alpha$

$$\mathcal{R}\left(\mathcal{E}^{\mathrm{R}}_{a_\bullet}, A\rho_\star\right) \ge 1-\alpha,$$

i.e. $\rho_\star^2 = \min_{k\in\mathbb{N}}\left\{a_k^2 \vee \frac{\nu_k^2}{n}\right\}$ is a lower bound for the minimax radius of testing.

Proof of Proposition 3.4.1. **Reduction step.** To prove a lower bound for the testing radius we reduce the risk of a test to a distance between probability measures. Denote $\mathbb{P}_0 := \mathbb{P}_{f^\circ}$ and let $\mathbb{P}_1$, specified below, be a mixing measure over the $\underline{A}_\alpha\rho_\star$-separated alternative. The minimax risk can then be lower bounded by applying a classical reduction argument as follows

$$\begin{aligned}\mathcal{R}\left(\mathcal{E}^{\mathrm{R}}_{a_\bullet}, \underline{A}_\alpha\rho_\star\right) &= \inf_{\Delta}\left\{\mathbb{P}_{f^\circ}(\Delta=1) + \sup_{f-f^\circ\in\mathscr{L}^2_{\underline{A}_\alpha\rho_\star}\cap\mathcal{E}, f\in\mathcal{D}} \mathbb{P}_f(\Delta=0)\right\}\\ &\ge \inf_{\Delta}\{\mathbb{P}_0(\Delta=1)+\mathbb{P}_1(\Delta=0)\}\\ &= 1-\mathrm{TV}(\mathbb{P}_0,\mathbb{P}_1) \ge 1-\sqrt{\frac{\chi^2(\mathbb{P}_0,\mathbb{P}_1)}{2}},\end{aligned} \tag{3.4.3}$$

where TV denotes the total variation distance and χ^2 the χ^2-divergence. The last inequality follows e.g. from Lemma 2.5 combined with (2.7) in Tsybakov [2009].

Definition of the mixture. On the alternative we mix the Fourier coefficients uniformly over the vertices of a hypercube. Consider f with $f - f^\circ \in \mathcal{E}^{\mathrm{R}}_{a_\bullet} \cap \mathscr{L}^2_{\underline{A}_\alpha \rho_\star}$ with coefficients

$$f_j := \begin{cases} 1 & j = 0, \\ \frac{\sqrt{\zeta\eta\rho_\star}}{\nu^2_{\kappa_\star}} |\varphi_j|^2 & |j| \in [\![\kappa_\star]\!], \\ 0 & \text{otherwise,} \end{cases}$$

where $\zeta := \mathrm{R}^2 \wedge \sqrt{\log(1+2\alpha^2)} \wedge \mathfrak{a}^{-1}$. For a sign vector $\tau \in \{\pm\}^{\kappa_\star}$ we define f^τ with $f^\tau - f^\circ \in \mathcal{E}^{\mathrm{R}}_{a_\bullet} \cap \mathscr{L}^2_{\underline{A}_\alpha \rho_\star}$ through its Fourier coefficients

$$f^\tau_j = \begin{cases} 1 & j = 0, \\ \tau_{|j|} f_j & |j| \in [\![\kappa_\star]\!], \\ 0 & \text{otherwise.} \end{cases}$$

The quadratic functionals $\mathrm{q}^2(f^\tau - f^\circ) = \mathrm{q}^2(f - f^\circ)$ and $\mathrm{q}^2_k(f^\tau - f^\circ) = \mathrm{q}^2_k(f - f^\circ)$, $k \in \mathbb{N}$ are invariant under τ. The resulting mixing measure is given by

$$\mathbb{P}_1 := \frac{1}{2^{\kappa_\star}} \sum_{\tau \in \{\pm\}^{\kappa_\star}} \mathbb{P}_{f^\tau}.$$

Let us check that the constructed candidates f^τ, $\tau \in \{\pm\}^{\kappa_\star}$ are indeed densities and are contained in the alternative. Let $\tau \in \{\pm\}^{\kappa_\star}$.

(a) $\sum_{j \in \mathbb{Z}} \left| f^\tau_j \right|^2 < \infty$ ($\in \mathscr{L}^2$)
Satisfied by construction.

(b) $f^\tau_j = \overline{f^\tau_{-j}}$ (real-valued)
Satisfied by construction.

(c) $f^\tau_0 = 1$ (normalized to 1)
Satisfied by construction.

(d) $\sum_{|j| \in \mathbb{N}} \left| f^\tau_j \right| \leq 1$ (positive)
The Cauchy-Schwarz inequality implies

$$\begin{aligned} \sum_{|j| \in \mathbb{N}} \left| f^\tau_j \right| &\leq \left(\sum_{|j| \in \mathbb{N}} a^2_{|j|} \right)^{1/2} \left(\sum_{|j| \in \mathbb{N}} a^{-2}_{|j|} |f_j|^2 \right)^{1/2} \\ &= \left(2 \sum_{j \in \mathbb{N}} a^2_j \right)^{1/2} \left(2 \sum_{j \in \mathbb{N}} a^{-2}_j |f_j|^2 \right)^{1/2} \leq \sqrt{\zeta}\sqrt{\mathfrak{a}} \leq 1, \end{aligned}$$

where the second last inequality follows as in (e).

(e) $f - f^\circ \in \mathcal{E}^{\mathrm{R}}_{a_\bullet}$, i.e. $2\sum_{j \in \mathbb{N}} a^{-2}_j |f_j|^2 \leq \mathrm{R}^2$ (smoothness)
By the monotonicity of $a_\bullet$ and the definition of ζ we have

$$\begin{aligned} 2 \sum_{j \in \mathbb{N}} a^{-2}_{|j|} |f_j|^2 &= \frac{\zeta\eta\rho^2_\star}{\nu^4_{\kappa_\star}} \sum_{|j| \in [\![\kappa_\star]\!]} |\varphi_j|^{-4} a^{-2}_{|j|} \\ &\leq \frac{\zeta\eta\rho^2_\star}{\nu^4_{\kappa_\star}} a^{-2}_{\kappa_\star} \nu^4_{\kappa_\star} \leq \zeta\eta\rho^2_\star a^{-2}_{\kappa_\star} \leq \zeta \leq \mathrm{R}^2. \end{aligned}$$

(f) $f - f^\circ \in \mathscr{L}^2_{\underline{A}_\alpha \rho_\star}$, i.e. $\mathrm{q}(f - f^\circ) \geq \underline{A}_\alpha \rho_\star$. (separation)
By definition

$$\mathrm{q}^2(f - f^\circ) = \mathrm{q}^2_{\kappa_\star}(f - f^\circ) = \frac{\zeta \eta \rho_\star^2}{\nu_{\kappa_\star}^4} \sum_{|j| \in [\![\kappa_\star]\!]} |\varphi_j|^{-4} = \zeta \eta \rho_\star^2 = \underline{A}_\alpha^2 \rho_\star^2.$$

(g) $n^2 \sum_{|j| \in [\![\kappa_\star]\!]} |f_j|^4 |\varphi_j|^4 \leq \log(1 + 2\alpha^2)$ (similarity)
The definition of ζ and the condition on η imply

$$\begin{aligned} n^2 \sum_{|j| \in [\![\kappa_\star]\!]} |f_j|^4 |\varphi_j|^4 &= n^2 \frac{\zeta^2 \eta^2 \rho_\star^4}{\nu_{\kappa_\star}^8} \sum_{|j| \in [\![\kappa_\star]\!]} |\varphi_j|^{-4} \\ &= n^2 \frac{\zeta^2 \eta^2 \rho_\star^4}{\nu_{\kappa_\star}^2} \leq n^2 \frac{\zeta^2 \eta^2 \rho_\star^4}{\nu_{\kappa_\star}^2} \leq \zeta^2 \leq \log(1 + 2\alpha^2). \end{aligned}$$

The conditions (a)-(d) guarantee that the vertices are densities, (e) and (f) guarantee that the vertices lie in the alternative. Condition (g) is needed to bound the χ^2-divergence between $\mathbb{P}_0$ and $\mathbb{P}_1$ and thus guarantees that the induced distance between the mixing measure and the null is negligible, i.e. that they are *similar* enough to be statistically indistinguishable.
Bound for the χ^2-divergence. We apply Lemma C.1.2 from the appendix and obtain

$$\chi^2 \left(\frac{1}{2^{\kappa_\star}} \sum_{\tau \in \{\pm\}^{\kappa_\star}} \mathbb{P}_{f^\tau}, \mathbb{P}_0 \right) \leq \exp \left(2n^2 \sum_{j \in [\![\kappa_\star]\!]} |g_j|^4 \right) - 1 = \exp \left(n^2 \sum_{|j| \in [\![\kappa_\star]\!]} |f_j|^4 |\varphi_j|^4 \right) - 1.$$

Hence, the condition (g) implies

$$\chi^2 \left(\frac{1}{2^{\kappa_\star}} \sum_{\tau \in \{\pm\}^{\kappa_\star}} \mathbb{P}_{f^\tau}, \mathbb{P}_0 \right) \leq \exp \left(\log(1 + 2\alpha^2) \right) - 1 = 2\alpha^2$$

and, therefore, by inserting this bound into (3.4.3)

$$\mathcal{R}\left(\mathcal{E}^{\mathrm{R}}_{a_\bullet}, \underline{A}_\alpha \rho_\star\right) \geq 1 - \sqrt{\frac{\chi^2(\mathbb{P}_0, \mathbb{P}_1)}{2}} \geq 1 - \alpha,$$

which proves the claim. □

Remark 3.4.2 (Conditions on η and $\mathfrak{a}$). *Proposition 3.4.1 involves the value η satisfying (3.4.2), which depends on the joint behaviour of the sequences $a_\bullet$ and $\varphi_\bullet$ and essentially guarantees an optimal balance of the bias and the variance term in the dimension $\kappa_\star$. For all typical smoothness and ill-posedness assumptions considered in Illustration 3.2.6 an η exists such that (3.4.2) holds uniformly over all $n \in \mathbb{N}$. The additional assumption $\mathfrak{a} = 2\sum_{j \in \mathbb{N}} a_j^2 < \infty$ in Proposition 3.4.1 is needed to ensure that the candidate densities constructed in the reduction scheme of the proof are indeed densities. This assumption is in particular satisfied for the typical smoothness classes introduced in Illustration 3.2.6. For Sobolev-type alternatives, i.e. $a_j \sim j^{-s}$ it is satisfied as soon as $s > 1/2$, for super smooth alternatives, i.e. $a_j \sim \exp(-j^s)$ it is satisfied for all positive s.* □

3.5 Connection between quadratic functional estimation and testing

In this section we explore the connection between quadratic functional estimation and testing. Every estimator for the quadratic functional $\mathrm{q}^2(f - f^\circ) = \|f - f^\circ\|^2_{\mathscr{L}^2}$ can be used to construct a test by rejecting the null as soon as the estimated value of the quadratic functional exceeds a certain threshold. The next proposition shows how this connection can be formalized in terms of the minimax risk and the minimax radius. Denote by $0_{\mathscr{L}^2}$ the null element in in $\mathscr{L}^2$. Note that since the non-parametric class $\mathcal{E}$ is formulated as a constraint on $f - f^\circ$, it is natural to assume $0_{\mathscr{L}^2} \in \mathcal{E}$.

Proposition 3.5.1 (Testing is faster than quadratic functional estimation). Let $\alpha \in (0,1)$, $\mathcal{E} \subseteq \mathscr{L}^2$ a nonparametric class with $0_{\mathscr{L}^2} \in \mathcal{E}$ and $\rho^2(\mathcal{E})$ a minimax radius of testing with $\underline{A}_\alpha$ as in the lower bound definition. Then, the minimax risk of estimation satisfies

$$r^2(\mathcal{E}) \geq (1-\alpha)\frac{\underline{A}_\alpha^2}{8} \cdot \rho^4(\mathcal{E}).$$

Proof of Proposition 3.5.1. Let $\hat{\mathrm{q}}^2$ be any estimator of $\mathrm{q}^2(f - f^\circ)$. Define the test $\Delta := \mathbb{1}_{\{\hat{\mathrm{q}}^2 \geq \rho^2/2\}}$ with $\rho = \underline{A}_\alpha \rho(\mathcal{E})$. We convert the mean squared error into the sum of type I and type II error probabilities, i.e. the testing risk, by applying Markov's inequality. Keeping in mind that $\mathrm{q}^2(f^\circ - f^\circ) = 0$, we have

$$\begin{aligned}
r^2(\hat{\mathrm{q}}^2, \mathcal{E}) &= \sup_{f - f^\circ \in \mathcal{E}} \mathbb{E}_f \left(\hat{\mathrm{q}}^2 - \mathrm{q}^2(f - f^\circ)\right)^2 \\
&\geq \frac{1}{2}\left\{\mathbb{E}_{f^\circ}\left(\hat{\mathrm{q}}^2 - \mathrm{q}^2(f^\circ - f^\circ)\right) + \sup_{f - f^\circ \in \mathcal{E} \cap \mathscr{L}^2_\rho} \mathbb{E}_f \left(\hat{\mathrm{q}}^2 - \mathrm{q}^2(f - f^\circ)\right)\right\} \\
&\geq \frac{\rho^4}{8}\left\{\mathbb{P}_{f^\circ}\left(\hat{\mathrm{q}}^2 \geq \frac{\rho^2}{2}\right) + \sup_{f - f^\circ \in \mathcal{E} \cap \mathscr{L}^2_\rho} \mathbb{P}_f\left(\mathrm{q}^2(f - f^\circ) - \hat{\mathrm{q}}^2 \geq \frac{\rho^2}{2}\right)\right\} \\
&\geq \frac{\rho^4}{8}\left\{\mathbb{P}_{f^\circ}\left(\hat{\mathrm{q}}^2 \geq \frac{\rho^2}{2}\right) + \sup_{f - f^\circ \in \mathcal{E} \cap \mathscr{L}^2_\rho} \mathbb{P}_f\left(\hat{\mathrm{q}}^2 \leq \frac{\rho^2}{2}\right)\right\} \\
&= \frac{\rho^4}{8}\mathcal{R}\left(\Delta \mid \mathcal{E}, \underline{A}_\alpha \rho(\mathcal{E})\right).
\end{aligned}$$

Since $\hat{\mathrm{q}}^2$ is arbitrary and by definition $\mathcal{R}\left(\mathcal{E}, \underline{A}_\alpha \rho(\mathcal{E})\right) \geq 1 - \alpha$, we obtain the result. □

3.6 Lower bound for the estimation risk

Recall that the upper bound for the risk of estimation in Theorem 3.2.4 is of order $\rho_\star^4 \vee r_\circ^4$. There are two possible scenarios, either the risk is governed by the **bias^2-variance-**term $\rho_\star^4 = \min_{k \in \mathbb{N}}\left\{a_k^4 \vee \frac{\nu_k^4}{n^2}\right\}$ or by the **base level** term $r_\circ^4 = \max_{m \in \mathbb{N}}\left\{a_m^4 \wedge \frac{a_m^2}{n|\varphi_m|^2}\right\}$. We prove separate lower bounds for these two cases. The lower bound in the first case is an immediate consequence of the lower bound for the radius of testing Proposition 3.4.1 combined with Proposition 3.5.1.

Corollary 3.6.1 (First lower bound for the risk of estimation). Let $\mathfrak{a} \in (0, \infty)$ and $\eta \in (0,1]$ satisfy (3.4.1) and (3.4.2), respectively. Then, for all $n \geq 2$ we have

$$r^2(\mathcal{E}^{\mathrm{R}}_{a_\bullet}) \geq \frac{\eta^2(\mathrm{R}^4 \wedge \log(3/2) \wedge \mathfrak{a}^{-1})}{16} \min_{k \in \mathbb{N}}\left\{a_k^4 \vee \frac{\nu_k^4}{n^2}\right\}.$$

Proof of Corollary 3.6.1. We apply Proposition 3.5.1 to the lower bound for the radius of testing derived in Proposition 3.4.1 and set $\alpha = 0.5$. □

Let us now turn to the second lower bound. In contrast to the lower bound proved in Proposition 3.4.1, the proof of the next proposition only requires the construction of two candidate densities.

Proposition 3.6.2 (Second lower bound for the risk of estimation). For all $n \geq 2$ we have

$$r^2(\mathcal{E}^{\mathrm{R}}_{a_\bullet}) \geq \left(\frac{1}{64} \wedge \frac{\mathrm{R}^4}{16}\right) \max_{m\in\mathbb{N}} \left\{ a_m^4 \wedge \frac{a_m^2}{n\,|\varphi_m|^2} \right\}.$$

Proof of Proposition 3.6.2. **Reduction step.** Denote by $\mathbb{Q}_f$ the measure with density $f \circledast \varphi$. The measure $\mathbb{P}_f$ associated with the observations equals the n-fold product measure of $\mathbb{Q}_f$. Let $f^+, f^- \in \mathcal{D}$ (to be specified below) with associated $\mathbb{P}_{f^+}$, $\mathbb{P}_{f^-}$ and quadratic functionals $\mathrm{q}^2 = \mathrm{q}^2(f^+)$ and $\mathrm{p}^2 = \mathrm{q}^2(f^-)$. Denote by $\mathrm{h}(\mathbb{P}_{f^+}, \mathbb{P}_{f^-})$ the Hellinger affinity between the two measures $\mathbb{P}_{f^+}$ and $\mathbb{P}_{f^-}$. We apply the reduction scheme of Lemma C.2.1 and obtain

$$r^2(\mathcal{E}^{\mathrm{R}}_{a_\bullet}) \geq \frac{\mathrm{h}^2(\mathbb{P}_+, \mathbb{P}_-)}{8} \left(\mathrm{q}^2 - \mathrm{p}^2\right)^2. \tag{3.6.1}$$

Using the tensorization property of the Hellinger affinity and the definition of the Hellinger distance (cp. for instance Tsybakov [2009], p. 83), it follows

$$\mathrm{h}(\mathbb{P}_{f^+}, \mathbb{P}_{f^-}) = \left(\mathrm{h}(\mathbb{Q}_{f^+}, \mathbb{Q}_{f^-})\right)^n = \left(1 - \frac{1}{2}\mathrm{H}^2(\mathbb{Q}_{f^+}, \mathbb{Q}_{f^-})\right)^n.$$

Let us denote $g^\pm := f^\pm \circledast \varphi$, we will ensure that $g^- \geq \frac{1}{2}$. Hence,

$$\mathrm{H}^2(\mathbb{Q}_{f^+}, \mathbb{Q}_{f^-}) = \int \frac{(g^+(x) - g^-(x))^2}{\left(\sqrt{g^+(x)} + \sqrt{g^-(x)}\right)^2} \mathrm{d}x \leq 2 \left\| g^+ - g^- \right\|^2_{\mathscr{L}^2}.$$

Moreover, we ensure that $\|g^+ - g^-\|^2_{\mathscr{L}^2} \leq 1$. Then Bernoulli's inequality ($(1+x)^r \geq 1 + rx$ for all $x \geq -1, r \geq 0$) implies

$$\mathrm{h}^2(\mathbb{P}_{f^-}, \mathbb{P}_{f^+}) \geq 1 - 2n \left\| g^+ - g^- \right\|^2_{\mathscr{L}^2}.$$

From (3.6.1) it follows

$$\begin{aligned} r^2(\mathcal{E}^{\mathrm{R}}_{a_\bullet}) &\geq \frac{1}{8} \left(\mathrm{q}^2 - \mathrm{p}^2\right)^2 \left(1 - 2n \left\| g^+ - g^- \right\|^2_{\mathscr{L}^2}\right) \\ &= \frac{1}{8} \left(\mathrm{q}^2 - \mathrm{p}^2\right)^2 \left(1 - 2n \left\| f^+ \circledast \varphi - f^- \circledast \varphi \right\|^2_{\mathscr{L}^2}\right). \end{aligned} \tag{3.6.2}$$

Construction of the hypotheses Let $\tau \in \{\pm\}$ and let m be arbitrary. Define the Fourier coefficients of the hypotheses f^τ, $\tau \in \{\pm\}$ by

$$f_j^+ = \begin{cases} 1 & j = 0, \\ (1+\xi) C a_m & j = \pm m, \\ 0 & \text{otherwise,} \end{cases} \qquad \text{and} \qquad f_j^- = \begin{cases} 1 & j = 0, \\ (1-\xi) C a_m & j = \pm m, \\ 0 & \text{otherwise,} \end{cases}$$

with $C := \frac{1}{4} \wedge \frac{\mathrm{R}}{\sqrt{8}}$ and $\xi^2 := 1 \wedge \frac{1}{n a_m^2 |\varphi_m|^2}$. Then, the hypotheses f^τ, $\tau \in \{\pm\}$ satisfy the following conditions:

1. $f^\tau \in \mathcal{D}$ (density)

(a) $\sum_{j\in\mathbb{Z}} \left|f_j^\tau\right|^2 < \infty$, by construction. $(\in \mathscr{L}^2)$

(b) $f_j^\tau = \overline{f_{-j}^\tau}$, by construction. (real-valued)

(c) $f_0^\tau = 1$, by construction. (normalized to 1)

(d) $\sum_{|j|\in\mathbb{N}} \left|f_j^\tau\right| = 2(1\pm\xi)Ca_m \leq 2\cdot 2Ca_m \leq 4C \leq 1$. (positive)

2. $f \circledast \varphi = g^- \geq \frac{1}{2}$ (bounded from below)

(e) $\sum_{|j|\in\mathbb{N}} \left|f_j^-\right| |\varphi_j| = 2(1-\xi)Ca_m |\varphi_m| \leq 2C \leq \frac{1}{2}$.

3. $f^\tau - f^\circ \in \mathcal{E}_{a_\bullet}^{\mathrm{R}}$ (smoothness)

(f) $2\sum_{j\in\mathbb{N}} a_j^{-2} \left|f_j^\tau\right|^2 = 2a_m^{-2}(1\pm\xi)^2 C^2 a_m^2 \leq 8C^2 \leq \mathrm{R}^2$.

4. $(\mathrm{p}^2 - \mathrm{q}^2)^2 \geq \left(\frac{1}{4} \wedge \mathrm{R}^4\right)\left\{a_m^4 \wedge \frac{a_m^2}{n|\varphi_m|^2}\right\}$ (separation)

(g) We have $\mathrm{q}^2(f^\tau - f^\circ) = \sum_{|j|\in\mathbb{N}} \left|f_j^\tau\right|^2 = 2(1\pm\xi)^2 a_m^2$, therefore,

$$
\begin{aligned}
(\mathrm{p}^2 - \mathrm{q}^2)^2 &= 4\left((1+\xi)^2 - (1-\xi)^2\right)^2 C^4 a_m^4 = 64\xi^2 C^4 a_m^4 \\
&= 64 \cdot \left(\frac{1}{4^4} \wedge \frac{\mathrm{R}^4}{8^2}\right) \xi^2 a_m^4 = \left(\frac{1}{4} \wedge \mathrm{R}^4\right) \xi^2 a_m^4 \\
&= \left(\frac{1}{4} \wedge \mathrm{R}^4\right)\left\{a_m^4 \wedge \frac{a_m^2}{n\,|\varphi_m|^2}\right\}.
\end{aligned}
$$

5. $\|f^+ \circledast \varphi - f^- \circledast \varphi\|_{\mathscr{L}^2}^2 \leq \frac{1}{4n}$ (similarity)

(h) We have $\|f^+ \circledast \varphi - f^- \circledast \varphi\|_{\mathscr{L}^2}^2 = 4C^2\xi^2 a_m^2 |\varphi_j|^2 \leq 4C^2 \frac{1}{n} \leq \frac{1}{4n}$.

Note that Condition (h) also implies $\|f^+ \circledast \varphi - f^- \circledast \varphi\|_{\mathscr{L}^2}^2 \leq 1$, which is a condition to apply Bernoulli's inequality. Combining both bounds (g) and (h) with the reduction in (3.6.2), we obtain

$$
r^2(\mathcal{E}_{a_\bullet}^{\mathrm{R}}) \geq \frac{1}{8}\left(\frac{1}{4} \wedge \mathrm{R}^4\right)\left\{a_m^4 \wedge \frac{a_m^2}{n\,|\varphi_m|^2}\right\}\left(1 - 2n\frac{1}{4n}\right) = \left(\frac{1}{64} \wedge \frac{\mathrm{R}^4}{16}\right)\left\{a_m^4 \wedge \frac{a_m^2}{n\,|\varphi_m|^2}\right\}.
$$

Since $m \in \mathbb{N}$ is arbitrary, this proves the assertion. □

3.7 Upper bound for the radius of testing via a direct test

This section is a first step towards an adaptive testing procedure, which we will explore in more detail in the next chapter. The test proposed in this section does not explicitly use the coefficients φ_j, $j \in \mathbb{Z}$ of the error density. Using a similar technique as in Section 3.3, we derive its radius of testing. We consider a test that is based on the estimation of

$$
\begin{aligned}
\mathrm{q}^2(g - g^\circ) = \mathrm{q}^2(f \circledast \varphi - f^\circ \circledast \varphi) &= \int_{[0,1)} (g(x) - g^\circ(x))^2\, \mathrm{d}y \\
&= \sum_{j\in\mathbb{Z}} \left|g_j - g_j^\circ\right|^2 = \sum_{j\in\mathbb{N}} \left|f_j - f_j^\circ\right|^2 |\varphi_j|^2,
\end{aligned}
$$

which we call the *direct* version of the quadratic functional, since we have *direct* access to observations of g. To be more precise, we consider

$$\hat{\mathrm{d}}_k^2 := 2 \sum_{|j| \in [\![k]\!]} \left\{ |\hat{g}_j|^2 - \frac{1 - |\hat{g}_j|^2}{n-1} \right\}, \tag{3.7.1}$$

where $\hat{g}_j := \frac{1}{n} \sum_{k=1}^{n} e_j(-Y_k)$ is the standard estimator for the jth Fourier coefficient g_j and $\frac{1-|\hat{g}_j|^2}{n-1}$ is a de-biasing term. Clearly, $\hat{\mathrm{d}}_k^2$ is an unbiased estimator of the truncated quantity $\mathrm{q}_k^2(g - g^\circ) = \sum_{|j| \in [\![k]\!]} \left| f_j - f_j^\circ \right|^2 |\varphi_j|^2$. The estimator can be written as a U-statistic

$$\hat{\mathrm{d}}_k^2 = \frac{1}{n(n-1)} \sum_{\substack{l \neq m \\ l,m \in [\![n]\!]}} \sum_{|j| \in [\![k]\!]} e_j(-Y_l) e_j(Y_m) =: \frac{1}{2} \mathrm{U}_n^{\mathrm{d}}$$

where $\mathrm{U}_n^{\mathrm{d}}$ is a canonical U-statistic

$$\mathrm{U}_n^{\mathrm{d}} = \binom{n}{2}^{-1} \sum_{\substack{l \neq m \\ l,m \in [\![n]\!]}} h(Y_l, Y_m)$$

with the symmetric kernel $h(y_1, y_2) := \sum_{|j| \in [\![k]\!]} e_j(-y_1) e_j(y_2)$. Let us first analyse the variance of the estimator (3.7.1).

Proposition 3.7.1 (Upper bound for the variance). Let $f^\circ = \mathbb{1}_{[0,1)}$ and $f \in \mathcal{D}$. For $n, k \in \mathbb{N}$, $n \geq 2$ the estimator defined in (3.7.1) satisfies

$$\mathrm{var}_f(\hat{\mathrm{d}}_k^2) \leq \|g\|_\infty \frac{1}{n} \sum_{|j| \in [\![k]\!]} |\varphi_j|^2 |f_j|^2 + \|g\|_\infty \frac{2k-1}{n^2}.$$

Proof of Proposition 3.7.1. Recall that $\hat{\mathrm{d}}_k^2 = \frac{1}{2} \mathrm{U}_n^{\mathrm{d}}$. Straight-forward calculations (similar to the proof of Proposition 3.2.1) show that the kernel h is symmetric and real-valued. Let us define the function

$$h_1 : [0,1) \longrightarrow \mathbb{R}, y \longmapsto h_1(y) := \mathbb{E}_f(h(y, Y_2)).$$

The variance of the U-statistic $\mathrm{U}_n^{\mathrm{d}}$ can thus be calculated using the formula (Lemma A on p. 183 in Serfling [2009])

$$\mathrm{var}_f(\mathrm{U}_n^{\mathrm{d}}) = \binom{n}{2}^{-1} (2(n-2)\xi_1 + \xi_2),$$

where $\xi_1 := \mathrm{var}_f(h_1(Y_1))$ and $\xi_2 := \mathrm{var}_f(h(Y_1, Y_2))$. Let us now find upper bounds for the quantities ξ_1 and ξ_2. We start with ξ_1. We have

$$h_1(y) = \mathbb{E}_f h(y, Y_2) = \sum_{|j| \in [\![k]\!]} e_j(-y) \mathbb{E}_f e_j(Y_2) = \sum_{|j| \in [\![k]\!]} e_j(-y) \overline{g_j},$$

i.e. the Fourier coefficients of h_1 are given by $h_{1,j} = \overline{g_{-j}}$, $|j| \in [\![k]\!]$ and zero otherwise. Hence, by Parseval's inequality we obtain

$$\xi_1 = \mathrm{var}_f(h_1(Y_1)) \leq \mathbb{E}_f |h_1(Y_1)|^2 \leq \|g\|_\infty \|h_1\|_2^2 = \|g\|_\infty \sum_{|j| \in [\![k]\!]} |g_j|^2 = \|g\|_\infty \sum_{|j| \in [\![k]\!]} |f_j|^2 |\varphi_j|^2.$$

For the term ξ_2 we note that

$$\operatorname{var}_f(h(Y_1, Y_2)) \leq \mathbb{E}_f |h(Y_1, Y_2)|^2 \leq \|g\|_\infty \int_{[0,1)} \int_{[0,1)} |h(y_1, y_2)|^2 \mathrm{d}y_1 g(y_2) \mathrm{d}y_2$$

with

$$\begin{aligned}\int_{[0,1)} |h(y_1, y_2)|^2 \mathrm{d}y_1 &= \sum_{|j|,|l| \in [\![k]\!]} \int e_j(-y_1) e_j(y_2) \overline{e_l(-y_1) e_l(y_2)} \mathrm{d}y_1 \\ &= \sum_{|j|,|l| \in [\![k]\!]} \int e_j(y_2 - y_1) \overline{e_j(y_2 - y_1)} \mathrm{d}y_1 = 2k - 1,\end{aligned}$$

where the last inequality follows from the orthonormality of the basis $(e_j)_{j \in \mathbb{Z}}$. Hence,

$$\xi_2 = \operatorname{var}_f(h(Y_1, Y_2)) \leq \|g\|_\infty (2k - 1).$$

Combining the bounds for ξ_1 and ξ_2 we obtain (using $\frac{1}{2(n-1)} \leq \frac{1}{n}$ for $n \geq 2$)

$$\begin{aligned}\operatorname{var}_f(\hat{\mathrm{d}}_k^2) = \frac{1}{4} \operatorname{var}_f(\mathrm{U}_n^{\mathrm{d}}) &\leq \frac{1}{n}\xi_1 + \frac{1}{n^2}\xi_2 \\ &\leq \|g\|_\infty \frac{1}{n} \sum_{|j| \in [\![k]\!]} |f_j|^2 |\varphi_j|^2 + \|g\|_\infty \frac{2k-1}{n^2},\end{aligned}$$

which proves the assertion. □

Since we want to analyse the behaviour of $\hat{\mathrm{d}}_k^2$ under the null hypotheses and the alternative, we state appropriate bounds for the variance in these two situations in the next corollary.

Corollary 3.7.2 (Variance upper bound under the null and the alternative). Let $f^\circ = \mathbb{1}_{[0,1)}$ and $f \in \mathcal{D}$. For $n, k \in \mathbb{N}$, $n \geq 2$ the estimator defined in (3.7.1) satisfies

$$\operatorname{var}_{f^\circ}(\hat{\mathrm{d}}_k^2) \leq \frac{2k-1}{n^2}, \tag{3.7.2}$$

$$\operatorname{var}_f(\hat{\mathrm{d}}_k^2) \leq \|\varphi\|_\infty \frac{\sqrt{2k-1}}{n} \mathrm{q}_k^2(g - g^\circ) + \|\varphi\|_\infty \frac{2k-1}{n^2}. \tag{3.7.3}$$

Proof of Corollary 3.7.2. We use the bound of Proposition 3.7.1. In particular, we have

$$\sum_{|j| \in [\![k]\!]} |\varphi_j|^2 |f_j|^2 = \mathrm{q}_k^2(g - g^\circ),$$

$1 \leq \sqrt{2k-1}$ and $\|g\|_\infty \leq \|\varphi\|_\infty$, which shows (3.7.3). For $f = f^\circ = \mathbb{1}_{[0,1)}$ we furthermore have $\mathrm{q}_k^2(g - g^\circ) = 0$ and $\|g\|_\infty \leq \sum_{|j| \in \mathbb{Z}} |f_j| |\varphi_j| \leq 1$, hence, the first assertion (3.7.2) follows. □

We consider a test that is based on the estimator (3.7.1) and compares it to a multiple of its standard deviation under the null. For $\alpha \in (0, 1)$, a constant C_α and $k \in \mathbb{N}$ define

$$\Delta_{k,\alpha}^{\mathrm{d}} := \mathbb{1}_{\left\{\hat{\mathrm{d}}_k^2 \geq C_\alpha \frac{\sqrt{2k-1}}{n}\right\}}. \tag{3.7.4}$$

Furthermore, for $k \in \mathbb{N}$ let

$$(\rho_k^{\mathrm{d}})^2 := a_k^2 \vee \left(\max_{|j| \in [\![k]\!]} |\varphi_j|^{-2}\right) \frac{\sqrt{2k-1}}{n}.$$

Proposition 3.7.3 (Upper bound for the radius of testing of $\Delta^{\mathrm{d}}_{k,\alpha/2}$). Let $\alpha \in (0,1)$ and let $C_{\alpha/2}$, $\tilde{A}_\alpha$ satisfy (3.3.1). Set $\overline{A}^2_\alpha := \mathrm{R}^2 + \tilde{A}^2_\alpha$. Then, for all $A \geq \overline{A}_\alpha$ and all $k \in \mathbb{N}$ we obtain

$$\mathcal{R}\left(\Delta^{\mathrm{d}}_{k,\alpha/2} \mid \mathcal{E}^{\mathrm{R}}_{a_\bullet}, A\rho^{\mathrm{d}}_k\right) \leq \alpha,$$

i.e. $(\rho^{\mathrm{d}}_k)^2$ is an upper bound for the radius of testing of $\left\{\Delta^{\mathrm{d}}_{k,\alpha/2}\right\}_{\alpha\in(0,1)}$.

Proof of Proposition 3.7.3. We show that both the type I error probability and the type II error probability of the test (3.7.4) are bounded by $\alpha/2$. For the **type I error probability** we apply Markov's inequality and obtain

$$\mathbb{P}_{f^\circ}\left(\Delta^{\mathrm{d}}_{k,\alpha/2} = 1\right) = \mathbb{P}_{f^\circ}\left(\hat{\mathrm{d}}^2_k \geq C_{\alpha/2}\frac{\sqrt{2k-1}}{n}\right) \leq \frac{\mathrm{var}_{f^\circ}(\hat{\mathrm{d}}^2_k)}{C^2_{\alpha/2}\frac{2k-1}{n^2}} \leq \frac{1}{C^2_{\alpha/2}},$$

where the last inequality follows from (3.7.2) in Corollary 3.7.2. Hence, the type I error probability is bounded by $\alpha/2$ for all $C_{\alpha/2} \geq \sqrt{2/\alpha}$. Since $\|\varphi\|_\infty > 1$, it in particular holds for all C_α satisfying (3.3.1). Next, we consider the **type II error probability**. Let f be contained in the $\overline{A}_\alpha\rho^{\mathrm{d}}_k$-separated alternative, i.e. $f - f^\circ \in \mathcal{E}^{\mathrm{R}}_{a_\bullet}$ and $\mathrm{q}^2(f-f^\circ) \geq \overline{A}^2_\alpha(\rho^{\mathrm{d}}_k)^2$. We expand inside of the type II error probability, centring the estimator $\hat{\mathrm{d}}^2_k$ by its expectation

$$\begin{aligned}\mathbb{P}_f(\Delta^{\mathrm{d}}_{k,\alpha/2} = 0) &= \mathbb{P}_f\left(\hat{\mathrm{d}}^2 < C_{\alpha/2}\frac{\sqrt{2k-1}}{n}\right)\\ &= \mathbb{P}_f\left(\hat{\mathrm{d}}^2 - \mathrm{q}^2_k(g-g^\circ) < C_{\alpha/2}\frac{\sqrt{2k-1}}{n} - \mathrm{q}^2_k(g-g^\circ)\right)\end{aligned}$$

and distinguish the following two cases

1. $\mathrm{q}^2_k(g-g^\circ) \geq 2C_{\alpha/2}\frac{\sqrt{2k-1}}{n}$, (easy to test)
2. $\mathrm{q}^2_k(g-g^\circ) < 2C_{\alpha/2}\frac{\sqrt{2k-1}}{n}$. (difficult to test)

Case 1: (easy to test) The case distinction implies $C_{\alpha/2}\frac{\sqrt{2k-1}}{n} - \mathrm{q}^2_k(g-g^\circ) \leq -\frac{1}{2}\mathrm{q}^2_k(g-g^\circ)$ and, therefore, we obtain by applying Markov's inequality

$$\begin{aligned}\mathbb{P}_f(\Delta^{\mathrm{d}}_{k,\alpha/2} = 0) &\leq \mathbb{P}_f\left(\hat{\mathrm{d}}^2_k - \mathrm{q}^2_k(g-g^\circ) \leq -\frac{1}{2}\mathrm{q}^2_k(g-g^\circ)\right)\\ &= \mathbb{P}_f\left(\mathrm{q}^2_k(g-g^\circ) - \hat{\mathrm{d}}^2_k \geq \frac{1}{2}\mathrm{q}^2_k(g-g^\circ)\right) \leq 4\cdot\frac{\mathrm{var}_f(\hat{\mathrm{d}}^2_k)}{(\mathrm{q}^2_k(g-g^\circ))^2}.\end{aligned}$$

Inserting the bound for the variance obtained in (3.7.3) of Corollary 3.7.2 and exploiting the case distinction condition yields

$$\mathbb{P}_f(\Delta^{\mathrm{d}}_{k,\alpha/2} = 0) \leq 4\cdot\|\varphi\|_\infty\left(\frac{\frac{\sqrt{2k-1}}{n}\mathrm{q}^2_k(g-g^\circ) + \frac{2k-1}{n^2}}{(\mathrm{q}^2_k(g-g^\circ))^2}\right) \leq \|\varphi\|_\infty\left(\frac{2}{C_{\alpha/2}} + \frac{1}{C^2_{\alpha/2}}\right) \leq \frac{\alpha}{2}.$$

Case 2: (difficult to test) Recall (cp. the proof of Proposition 3.3.1) that under the alternative we have

$$\mathrm{q}^2(f-f^\circ) = \sum_{|j|\in\mathbb{N}}\left|f_j - f^\circ_j\right|^2 \geq \overline{A}^2_\alpha(\rho^{\mathrm{d}}_k)^2 \qquad \text{and} \qquad \sum_{|j|>k}\left|f_j - f^\circ_j\right|^2 \leq a^2_k\mathrm{R}^2.$$

Hence,

$$\mathrm{q}_k^2(f-f^\circ) = \mathrm{q}^2(f-f^\circ) - \sum_{|j|>k} \left| f_j - f_j^\circ \right|^2 \geq \overline{A}_\alpha^2(\rho^{\mathrm{d}})^2 - \mathrm{R}^2 a_k^2$$
$$\geq \widetilde{A}_\alpha^2 (\max_{|j|\in[\![k]\!]} |\varphi_j|^{-2}) \frac{\sqrt{2k-1}}{n} + \mathrm{R}^2 a_k^2 - \mathrm{R}^2 a_k^2 = \widetilde{A}_\alpha^2 (\max_{|j|\in[\![k]\!]} |\varphi_j|^{-2}) \frac{\sqrt{2k-1}}{n}.$$

Moreover, we have the following connection between the *indirect* and the *direct* quadratic functionals

$$(\max_{|j|\in[\![k]\!]} |\varphi_j|^{-2}) \mathrm{q}_k^2(g-g^\circ) = (\max_{|j|\in[\![k]\!]} |\varphi_j|^{-2}) \sum_{|j|\in[\![k]\!]} \left| f_j - f_j^\circ \right|^2 |\varphi_j|^2$$
$$\geq (\max_{|j|\in[\![k]\!]} |\varphi_j|^{-2})(\min_{|j|\in[\![k]\!]} |\varphi_j|^2) \sum_{|j|\in[\![k]\!]} \left| f_j - f_j^\circ \right|^2 = \mathrm{q}_k^2(f-f^\circ).$$

Therefore, under the alternative $\mathrm{q}_k^2(g-g^\circ) \geq \widetilde{A}_\alpha^2 \frac{\sqrt{2k-1}}{n}$. and the type II error probability satisfies

$$\mathbb{P}_f(\Delta^{\mathrm{d}}_{k,\alpha/2} = 0) = \mathbb{P}_f\left(\hat{\mathrm{d}}_k^2 - \mathrm{q}_k^2(g-g^\circ) \leq C_{\alpha/2} \frac{\sqrt{2k-1}}{n} - \mathrm{q}_k^2(g-g^\circ) \right)$$
$$\leq \mathbb{P}_f\left(\hat{\mathrm{d}}_k^2 - \mathrm{q}_k^2(g-g^\circ) \leq (C_{\alpha/2} - \widetilde{A}_\alpha^2) \frac{\sqrt{2k-1}}{n} \right)$$
$$= \mathbb{P}_f\left(\mathrm{q}_k^2(g-g^\circ) - \hat{\mathrm{d}}_k^2 \geq (\widetilde{A}_\alpha^2 - C_{\alpha/2}) \frac{\sqrt{2k-1}}{n} \right) \leq \frac{\mathrm{var}_f(\hat{\mathrm{d}}_k^2)}{(\widetilde{A}_\alpha^2 - C_{\alpha/2})^2 \frac{2k-1}{n^2}}.$$

Exploiting the bound for the variance (3.7.3) of Corollary 3.7.2, the case distinction condition and the choice of $\widetilde{A}_\alpha$ and $C_{\alpha/2}$, it follows

$$\mathbb{P}_f(\Delta^{\mathrm{d}}_{k,\alpha/2} = 0) \leq \|\varphi\|_\infty \cdot \frac{2C_{\alpha/2}+1}{\left(\widetilde{A}_\alpha^2 - C_{\alpha/2}\right)^2} \leq \frac{\alpha}{2},$$

which completes the proof. □

Define the minimum and minimizer that realize the bias-variance trade-off by

$$(\rho_\star^{\mathrm{d}})^2 := \min_{k\in\mathbb{N}} \left\{ a_k^2 \vee (\max_{|j|\in[\![k]\!]} |\varphi_j|^{-2}) \frac{\sqrt{2k-1}}{n} \right\},$$
$$\kappa_\star^{\mathrm{d}} := \arg\min_{k\in\mathbb{N}} \left\{ a_k^2 \vee (\max_{|j|\in[\![k]\!]} |\varphi_j|^{-2}) \frac{\sqrt{2k-1}}{n} \right\}.$$

Then Proposition 3.7.3 immediately yields the following corollary.

Corollary 3.7.4 (Optimised upper bound for the radius of testing of $\Delta^{\mathrm{d}}_{\kappa_\star^{\mathrm{d}},\alpha/2}$). Let $\alpha \in (0,1)$ and let $C_{\alpha/2}$, $\tilde{A}_\alpha$ satisfy (3.3.1). Set $\overline{A}_\alpha^2 := \mathrm{R}^2 + \widetilde{A}_\alpha^2$. Then, for all $A \geq \overline{A}_\alpha$ and all $k \in \mathbb{N}$ we obtain

$$\mathcal{R}\left(\Delta^{\mathrm{d}}_{\kappa_\star,\alpha/2} \mid \mathcal{E}^{\mathrm{R}}_{a_\bullet}, A\rho_\star^{\mathrm{d}} \right) \leq \alpha,$$

i.e. $(\rho_\star^{\mathrm{d}})^2$ is an upper bound for the radius of testing of $\left\{ \Delta^{\mathrm{d}}_{\kappa_\star^{\mathrm{d}},\alpha/2} \right\}_{\alpha\in(0,1)}$.

Appendix C

Auxiliary results

C.1 Auxiliary results for proving lower bounds of testing

Lemma C.1.1 (Interchanging sums and products on cubes). Let $k \in \mathbb{N}$. For sign vectors $\tau \in \{\pm\}^k$ we introduce $J^\tau = (J_j^{\tau_j})_{j \in [\![k]\!]} \subseteq \mathbb{R}^k$. Then,

$$\frac{1}{2^k} \sum_{\tau \in \{\pm\}^k} \prod_{j \in [\![k]\!]} J_j^{\tau_j} = \prod_{j \in [\![k]\!]} \frac{J_j^- + J_j^+}{2}.$$

Proof of Lemma C.1.1. The proof is by induction over k.
The **base case** $k = 1$ follows immediately, since

$$\frac{1}{2} \sum_{\tau_1 \in \{\pm\}} J_1^{\tau_1} = \frac{J_1^- + J_1^+}{2}.$$

For the **induction step** assume

$$\frac{1}{2^k} \sum_{\tau \in \{\pm\}^k} \prod_{j \in [\![k]\!]} J_j^{\tau_j} = \prod_{j \in [\![k]\!]} \frac{J_j^- + J_j^+}{2}.$$

Then it follows,

$$\begin{aligned}
&\frac{1}{2^{k+1}} \sum_{\tau \in \{\pm\}^{k+1}} \prod_{j \in [\![k+1]\!]} J_j^{\tau_j} \\
&= \frac{1}{2^{k+1}} \left(\left(\sum_{\substack{\tau \in \{\pm\}^{k+1} \\ \tau_{k+1} = +}} \prod_{j \in [\![k+1]\!]} J_j^{\tau_j} \right) + \left(\sum_{\substack{\tau \in \{\pm\}^{k+1} \\ \tau_{k+1} = -}} \prod_{j \in [\![k+1]\!]} J_j^{\tau_j} \right) \right) \\
&= \frac{1}{2^{k+1}} \left(\left(\sum_{\tau \in \{\pm\}^k} \prod_{j \in [\![k]\!]} J_j^{\tau_j} \right) J_{k+1}^+ + \left(\sum_{\tau \in \{\pm\}^k} \prod_{j \in [\![k]\!]} J_j^{\tau_j} \right) J_{k+1}^- \right) \\
&= \frac{1}{2} \left(J_{k+1}^+ + J_{k+1}^- \right) \left(\frac{1}{2^k} \sum_{\tau \in \{\pm\}^k} \prod_{j \in [\![k]\!]} J_j^{\tau_j} \right) \\
&= \frac{1}{2} \left(J_{k+1}^+ + J_{k+1}^- \right) \prod_{j \in [\![k]\!]} \frac{J_j^+ + J_j^-}{2} = \prod_{j \in [\![k+1]\!]} \frac{J_j^+ + J_j^-}{2},
\end{aligned}$$

where the induction assumption is used in the second last step. □

Lemma C.1.2 (χ^2-divergence for mixtures over hypercubes). Let $k \in \mathbb{N}$. For $\tau \in \{\pm\}^k$ define coefficients $\theta^\tau_\bullet \in \ell^2(\mathbb{Z})$ and functions $g^\tau \in \mathscr{L}^2$ by setting

$$\theta^\tau_j = \begin{cases} \tau_{|j|}\theta_{|j|} & |j| \in [\![k]\!], \\ 1 & j = 0, \\ 0 & |j| > k, \end{cases} \qquad \text{and} \qquad g^\tau = \sum_{j=-k}^{k} \theta^\tau_j e_j = e_0 + \sum_{|j|\in[\![k]\!]} \theta^\tau_j e_j.$$

Assuming $g^\tau \in \mathcal{D}$ for each $\tau \in \{\pm\}^k$, we consider the mixture $\mathbb{P}_1$ with probability density

$$\frac{1}{2^{k^s}} \sum_{\tau\in\{\pm\}^{k^s}} \prod_{i\in[\![n]\!]} g^\tau(z_i), \qquad \text{for } z_i \in [0,1), i \in [\![n]\!]$$

and denote $\mathbb{P}_0 = \mathbb{P}_{f^\circ}$ with probability density

$$\prod_{i\in[\![n]\!]} \mathbb{1}_{[0,1)}(z_i), \qquad \text{for } z_i \in [0,1), i \in [\![n]\!].$$

Then, the χ^2-divergence satisfies

$$\chi^2(\mathbb{P}_1, \mathbb{P}_2) \le \exp\left(2n^2 \sum_{j\in[\![k]\!]} \theta_j^4\right) - 1.$$

Proof of Lemma C.1.2. Since $\mathbb{P}_1 \ll \mathbb{P}_0$ we have

$$\chi^2(\mathbb{P}_1, \mathbb{P}_0) = \mathbb{E}_0\left(\frac{\mathrm{d}\mathbb{P}_1}{\mathrm{d}\mathbb{P}_0}(Z_1,\dots,Z_n)\right)^2 - 1.$$

for i.i.d. random variables $(Z_j)_{j\in[\![n]\!]}$ with marginal density $f^\circ = \mathbb{1}_{[0,1)}$ under $\mathbb{P}_0$. Let $z_j \in [0,1)$, $j \in [\![n]\!]$, then the likelihood ratio becomes

$$\frac{\mathrm{d}\mathbb{P}_1}{\mathrm{d}\mathbb{P}_0}(z_1,\dots,z_n) = \frac{1}{2^k} \sum_{\tau\in\{\pm\}^k} \prod_{i\in[\![n]\!]} g^\tau(z_i),$$

since $\mathbb{P}_0$ is a product over uniform densities. Squaring, taking the expectation under $\mathbb{P}_0$ and exploiting the independence yields

$$\begin{aligned}\mathbb{E}_0\left(\frac{\mathrm{d}\mathbb{P}_1}{\mathrm{d}\mathbb{P}_0}(Z_1,\dots,Z_n)\right)^2 &= \left(\frac{1}{2^k}\right)^2 \sum_{\tau,\eta\in\{\pm\}^k} \prod_{i\in[\![n]\!]} \mathbb{E}_0(g^\tau(Z_j)g^\eta(Z_j)) \\ &= \left(\frac{1}{2^k}\right)^2 \sum_{\tau,\eta\in\{\pm\}^k} \left(\mathbb{E}_0(g^\tau(Z_1)g^\eta(Z_1))\right)^n.\end{aligned}$$

Let us calculate

$$\mathbb{E}_0(g^\tau(Z_1)g^\eta(Z_1)) = \int_{[0,1)} g^\tau(z)g^\eta(z)\mathrm{d}z = 1 + 2\sum_{|j|\in[\![k]\!]} \theta^\tau_j\theta^\eta_j,$$

where the last equality is due to the orthonormality of $(e_j)_{j\in\mathbb{Z}}$ and the symmetry of θ^τ and θ^η. Applying the inequality $1 + x \le \exp(x)$, which holds for all $x \in \mathbb{R}$, we obtain

$$\mathbb{E}_0(g^\tau(Z_1)g^\eta(Z_1)) = 1 + 2\sum_{|j|\in[\![k]\!]} \theta^\tau_j\theta^\eta_j \le \exp(2\sum_{j\in[\![k]\!]} \theta^\tau_j\theta^\eta_j) = \prod_{j\in[\![k]\!]} \exp(2\theta^\tau_j\theta^\eta_j).$$

Hence,

$$\mathbb{E}_0\left(\frac{\mathrm{d}\mathbb{P}_1}{\mathrm{d}\mathbb{P}_0}(Z_1,\dots,Z_n)\right)^2 \le \left(\frac{1}{2^k}\right)^2 \sum_{\tau,\eta\in\{\pm\}^k}\prod_{j\in[\![k]\!]}\exp(2n\theta_j^{\tau}\theta_j^{\eta}),$$

where we can apply Lemma C.1.1 to the η-summation with $J_j^{\eta_j}=\exp(2n\theta_j^{\tau}\theta_j^{\eta})$ and obtain

$$\mathbb{E}_0\left(\frac{\mathrm{d}\mathbb{P}_1}{\mathrm{d}\mathbb{P}_0}(Z_1,\dots,Z_n)\right)^2 \le \frac{1}{2^k}\sum_{\tau\in\{\pm\}^k}\prod_{j\in[\![k]\!]}\frac{\exp(2n\theta_j^{\tau}\theta_j)+\exp(-2n\theta_j^{\tau}\theta_j)}{2},.$$

Applying Lemma C.1.1 to the τ-summation with $J_j^{\tau_j}=\frac{\exp(2n\theta_j^{\tau}\theta_j)+\exp(-2n\theta_j^{\tau}\theta_j)}{2}$ yields

$$\begin{aligned}\mathbb{E}_0\left(\frac{\mathrm{d}\mathbb{P}_1}{\mathrm{d}\mathbb{P}_0}(Z_1,\dots,Z_n)\right)^2 &\le \prod_{j\in[\![k]\!]}\frac{\exp(2n\theta_j^2)+\exp(-2n\theta_j^2)+\exp(-2n\theta_j^2)+\exp(2n\theta_j^2)}{4}\\ &= \prod_{j\in[\![k]\!]}\frac{\exp(2n\theta_j^2)+\exp(-2n\theta_j^2)}{2} = \prod_{j\in[\![k]\!]}\cosh(2n\theta_j^2).\end{aligned}$$

Since $\cosh(x)\le\exp(x^2/2)$ we obtain

$$\mathbb{E}_0\left(\frac{\mathrm{d}\mathbb{P}_1}{\mathrm{d}\mathbb{P}_0}(Z_1,\dots,Z_n)\right)^2 \le \prod_{j\in[\![k]\!]}\exp(2n^2\theta_j^4) = \exp\left(2n^2\sum_{j\in[\![k]\!]}\theta_j^4\right),$$

which completes the proof. □

C.2 Auxiliary results for proving lower bounds of estimation

Lemma C.2.1 (Reduction scheme for the estimation risk). Let $\mathcal{E}$ be a regularity class. For densities $f^+,f^-\in\mathscr{L}^2$ with $f^+-f^\circ, f^--f^\circ\in\mathcal{E}$ we have

$$\inf_{\hat{\mathrm{q}}^2}\sup_{f-f^\circ\in\mathcal{E}}\mathbb{E}_f\left(\hat{\mathrm{q}}^2-\mathrm{q}^2(f-f^\circ)\right)^2 \ge \frac{1}{8}\mathrm{h}^2(\mathbb{P}_{f^+},\mathbb{P}_{f^-})\left(\mathrm{q}^2(f^+-f^\circ)-\mathrm{q}^2(f^--f^\circ)\right)^2,$$

where $\mathrm{h}(\mathbb{P}_{f^+},\mathbb{P}_{f^-})$ denotes the Hellinger affinity between $\mathbb{P}_{f^+}$ and $\mathbb{P}_{f^-}$.

Proof of Lemma C.2.1. Let $\hat{\mathrm{q}}^2$ be any estimator and denote $\mathbb{P}_+ := \mathbb{P}_{f^+}$, $\mathbb{P}_- = \mathbb{P}_{f^-}$ and $\mathrm{q}^2 := \mathrm{q}^2(f^+-f^\circ)$ and $\mathrm{p}^2 := \mathrm{q}^2(f^--f^\circ)$. We have

$$\begin{aligned}\mathrm{h}(\mathbb{P}_+,\mathbb{P}_-) &= \int\sqrt{\mathrm{d}\mathbb{P}_+\mathrm{d}\mathbb{P}_-} = \int\left|\frac{\mathrm{q}^2-\mathrm{p}^2}{\mathrm{q}^2-\mathrm{p}^2}\right|\sqrt{\mathrm{d}\mathbb{P}_+\mathrm{d}\mathbb{P}_-}\\ &\le \int\left|\frac{\mathrm{q}^2-\hat{\mathrm{q}}^2}{\mathrm{q}^2-\mathrm{p}^2}\right|\sqrt{\mathrm{d}\mathbb{P}_+\mathrm{d}\mathbb{P}_-} + \int\left|\frac{\hat{\mathrm{q}}^2-\mathrm{p}^2}{\mathrm{q}^2-\mathrm{p}^2}\right|\sqrt{\mathrm{d}\mathbb{P}_+\mathrm{d}\mathbb{P}_-}\\ &\le \left(\int\left|\frac{\mathrm{q}^2-\hat{\mathrm{q}}^2}{\mathrm{q}^2-\mathrm{p}^2}\right|^2\mathrm{d}\mathbb{P}_+\right)^{1/2}\left(\int\mathrm{d}\mathbb{P}_-\right)^{1/2} + \left(\int\left|\frac{\hat{\mathrm{q}}^2-\mathrm{p}^2}{\mathrm{q}^2-\mathrm{p}^2}\right|^2\mathrm{d}\mathbb{P}_-\right)^{1/2}\left(\int\mathrm{d}\mathbb{P}_+\right)^{1/2}\\ &\le \frac{1}{|\mathrm{q}^2-\mathrm{p}^2|}\left(\left(\mathbb{E}_{f^+}(\mathrm{q}^2-\hat{\mathrm{q}}^2)^2\right)^{1/2}+\left(\mathbb{E}_{f^-}(\mathrm{p}^2-\hat{\mathrm{q}}^2)^2\right)^{1/2}\right)\\ &\le \frac{2}{|\mathrm{q}^2-\mathrm{p}^2|}\left(\mathbb{E}_{f^+}(\mathrm{q}^2-\hat{\mathrm{q}}^2)^2+\mathbb{E}_{f^-}(\mathrm{p}^2-\hat{\mathrm{q}}^2)^2\right)^{1/2}.\end{aligned}$$

Therefore,

$$\sup_{f-f^\circ\in\mathcal{E}} \mathbb{E}_f \left(\hat{\mathrm{q}}^2 - \mathrm{q}^2(f-f^\circ)\right)^2 \geq \frac{1}{2}\left(\mathbb{E}_{f^+}(\mathbb{q}^2-\hat{\mathrm{q}}^2)^2 + \mathbb{E}_{f^-}(\mathbb{p}^2-\hat{\mathrm{q}}^2)^2\right)$$
$$\geq \frac{\mathrm{h}^2(\mathbb{P}_+,\mathbb{P}_-)}{8}\left(\mathbb{q}^2-\mathbb{p}^2\right)^2,$$

which completes the proof. □

Chapter 4

Adaptive minimax testing for circular convolution

Given observations from a circular random variable contaminated by an additive measurement error, we consider the problem of minimax optimal goodness-of-fit testing in a non-asymptotic framework. We propose direct and indirect testing procedures using a projection approach. The structure of the optimal tests depends on regularity and ill-posedness parameters of the model, which are unknown in practice. Therefore, adaptive testing strategies that perform optimally over a wide range of regularity and ill-posedness classes simultaneously are investigated. Considering a multiple testing procedure, we obtain adaptive i.e. assumption-free procedures and analyse their performance. Compared with the non-adaptive tests, their radii of testing face a deterioration by a log-factor. We show that for testing of uniformity this loss is unavoidable by providing a lower bound. The results are illustrated considering Sobolev spaces and ordinary or super smooth error densities.

4.1 Adaptive testing

Minimax radii of testing in the circular model were derived in the previous chapter. Therein, we consider a test that is based on a projection estimator of the quantity $\|f - f^\circ\|^2_{\mathscr{L}^2}$, which depends on a dimension parameter. Choosing the dimension parameter optimally, the test that we obtain is shown to be minimax optimal, i.e. it achieves the minimax radius of testing (defined in Section 3.1.2) given by a typical bias^2-variance trade-off. The choice of the dimension parameter, however, depends on the underlying smoothness structure. Moreover, the test explicitly uses the coefficients of the error density. Since both are typically unknown in practise, we investigate adaptive testing strategies in this chapter, which do not rely on this prior knowledge.

Direct vs. indirect testing procedures We point out that estimating the energy $\|f - f^\circ\|^2_{\mathscr{L}^2}$ based on i.i.d. copies of $Y = X + \varepsilon - \lfloor X + \varepsilon \rfloor$ with density $g = f \circledast \varphi$ is an inverse problem, since it requires an inversion of the convolution transformation. This inversion introduces additional instability in deconvolution problems, caused by its ill-posedness. To circumvent this problem, in an inverse Gaussian sequence space model Laurent et al. [2011] argue for a **direct** testing procedure, which is based on the estimation of the energy in the image space of the operator. Let us explain this idea in our setting. Instead of the **indirect testing task**

$$H_0 : f - f^\circ = 0 \qquad \text{against} \qquad H_1 : f - f^\circ \neq 0,$$

which is called *indirect* since we do not have access to observations from the density f, we examine the **direct testing task**

$$H_0^{\mathrm{D}} : g - g^\circ = (f - f^\circ) \circledast \varphi = 0 \qquad \text{against} \qquad H_1^{\mathrm{D}} : g - g^\circ \neq 0,$$

where we have *direct* access to observations from g. This approach has two advantages: on the one hand the additional uncertainty caused by an inversion is avoided, on the other hand - in the special case $f^\circ = \mathbb{1}_{[0,1)}$ (and, hence, $g^\circ = f^\circ \circledast \varphi = \mathbb{1}_{[0,1)}$) - the proposed tests for the direct testing task no longer explicitly depend on the error density φ. For both the direct and indirect testing procedure the radii of testing, depending on a dimension parameter, are essentially determined by a bias^2-variance trade-off. As usual the optimal choice of the dimension parameter depends on both the smoothness of the alternative and on the ill-posedness of the model, which are unknown in practice. This motivates the study of adaptive testing procedures, which we investigate in this chapter.

Adaptive testing and related literature. In the literature **adaptive**, i.e. **assumption-free**, testing strategies have been studied in both an asymptotic and non-asymptotic framework. In an asymptotic framework e.g. Spokoiny [1996] considers adaptive testing in a sequence space model with Besov-type alternatives, showing that asymptotic adaptation comes with an unavoidable cost of a $\log\log$-factor. In a nonasymptotic setting Laurent et al. [2003] consider adaptive testing in a Gaussian regression model, Fromont and Laurent [2006] deal with a density model. Butucea [2007] and Butucea et al. [2009] determine adaptive rates of testing in a convolution model on the real line using kernel estimators of the $\mathscr{L}^2$-distance to the null. The proposed tests have as a common feature that they are based on estimators of the distance to the null, which only depend on the (unknown) smoothness through a tuning parameter (e.g. a bandwidth, a threshold or a dimension parameter). By aggregating the estimators over different tuning parameters into one test statistic – i.e. using a multiple testing approach – the authors obtain tests, which perform optimally over a wide range of alternatives. Since they no longer depend on the unknown regularity of the alternative, they are **assumption-free**. To formalise this idea, let us introduce a collection $\mathcal{A}$ of regularity parameters that characterise a family of alternatives $\{\mathcal{E}_{a_\bullet} : a_\bullet \in \mathcal{A}\}$ with corresponding radii $\{\rho_{a_\bullet}(n) := \rho(\mathcal{E}_{a_\bullet}) : a_\bullet \in \mathcal{A}\}$, where we now explicitly emphasise the dependence on the regularity parameter $a_\bullet \in \mathcal{A}$ and the number of observations n in the notation. In general, adaptation without a loss is impossible (cp. Spokoiny [1996]). To characterise the cost to pay for adaptation we introduce the **effective sample size** δn with $\delta = \delta_n$ depending on n. The factor $\delta \in [0, 1]$ shrinks the sample size n and, hence, evaluating the radius at δn deteriorates the radius of testing. In fact, the value δ^{-1} is called **adaptive factor** for the family of tests $\{\Delta_\alpha : \alpha \in (0, 1)\}$ over the family of alternatives $\{\mathcal{E}_{a_\bullet} : a_\bullet \in \mathcal{A}\}$, if for all $\alpha \in (0, 1)$ there exists a constant $\overline{A}_\alpha > 0$ such that

(i) for all $A \geq \overline{A}_\alpha$ we have $\sup_{a_\bullet \in \mathcal{A}} \mathcal{R}\left(\Delta_\alpha \mid \mathcal{E}, A\rho_{a_\bullet}(\delta n)\right) \leq \alpha$, (upper bound)

where $\rho_{a_\bullet}(n)$ denotes a radius of testing for the family $\{\Delta_\alpha : \alpha \in (0, 1)\}$. We shall emphasise that the testing risk now has to be bounded uniformly for all alternatives $\mathcal{E}_{a_\bullet}, a_\bullet \in \mathcal{A}$. We call δ^{-1} **minimal adaptive factor** if in addition for all $\alpha \in (0, 1)$ there exists a constant $\underline{A}_\alpha > 0$ such that

(ii) for all $A \leq \underline{A}_\alpha$ we have $\inf_\Delta \sup_{a_\bullet \in \mathcal{A}} \mathcal{R}\left(\Delta \mid \mathcal{E}, A\rho_{a_\bullet}(\delta n)\right) \geq 1 - \alpha$. (lower bound)

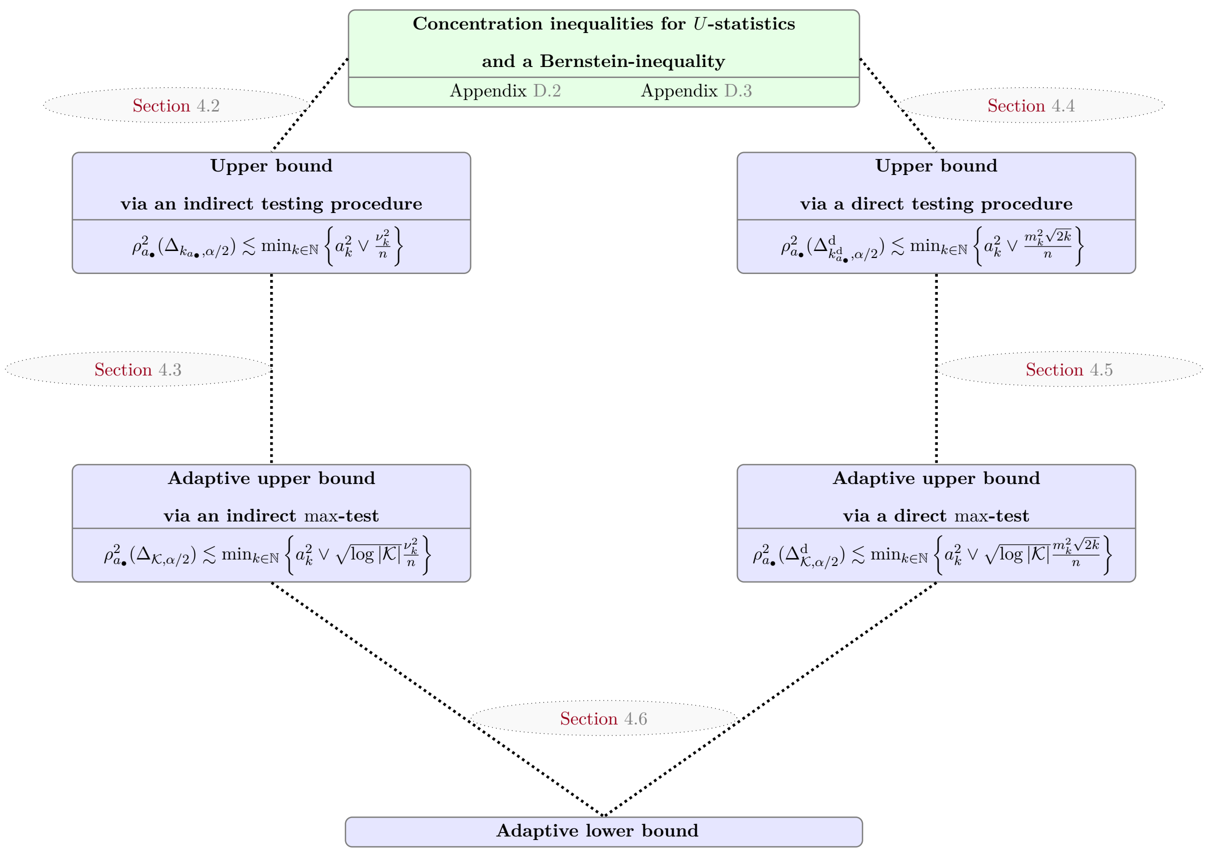
The goal of this chapter is to characterise the minimal adaptive factor δ^{-1}.

Aggregation procedure. Let us come back to the circular deconvolution problem and the indirect and direct tests discussed above. In this chapter we aggregate both testing procedures over a family $\mathcal{K} \subseteq \mathbb{N}$ of dimension parameters using a classical Bonferroni method, where for a given level $\alpha \in (0,1)$ each of the tests in the family has level $\frac{\alpha}{|\mathcal{K}|}$. The aggregated testing procedure rejects the null hypothesis as soon as one test in the collection rejects. It is straightforward to see that a Bonferroni aggregation of the direct and indirect tests proposed in the previous chapter leads to an adaptive factor of order $|\mathcal{K}|$ (since $C_\alpha \approx \frac{1}{\alpha}$ and, hence, $C_{\alpha/|\mathcal{K}|} \approx \frac{|\mathcal{K}|}{\alpha}$). The choice of the family $\mathcal{K}$ reflects the collection of alternatives over which the aggregated test performs optimally. If the alternatives characterise ordinary smoothness of the circular density, the size of $\mathcal{K}$ is typically chosen to be of order $\log n$ (cp. Fromont and Laurent [2006], Spokoiny [1996]). Then the aggregated test (from the previous chapter) will feature a deterioration by an adaptive factor of order $\log n$. However, we show in this chapter that generally the minimal adaptive factor is smaller. In order to do so, we first derive sharper bounds for the quantiles of the direct and indirect test statistics using exponential bounds for U-statistics and a Bernstein inequality (instead of the Markov inequality in the previous chapter). This allows to define a new version of an indirect and a direct test, for which we derive radii of testing. Aggregating these tests via the Bonferroni method we obtain an adaptive factor for adaptation with respect to smoothness of order $\sqrt{\log \log n}$. Interestingly, in case of testing for uniformity, i.e. $f^\circ = \mathbb{1}_{[0,1)}$, the aggregated direct test no longer depends on the noise density φ and is, thus, also adaptive with respect to the ill-posedness of the model. Moreover, in this situation we derive a lower bound for the adaptive factor providing conditions under which it is minimal.

Outline of this chapter. The upper bounds for the radius of testing via an indirect and a direct testing procedure are derived in Section 4.2 and Section 4.4, respectively. Section 4.3 and Section 4.5 are devoted to adaptive indirect and direct testing strategies. We provide lower bounds in Section 4.6.

Outline

Adaptive minimax testing for circular convolution



4.2 Upper bound via an indirect testing procedure

Notation and preliminaries. We consider the Hilbert space $\mathscr{L}^2 := \mathscr{L}^2[0,1)$ of square-integrable complex-valued functions defined on $[0,1)$ equipped with its usual norm $\|\cdot\|_{\mathscr{L}^2}$ and corresponding inner product $\langle f, g\rangle_{\mathscr{L}^2} := \int_0^1 f(x)\overline{g(x)}\mathrm{d}x$ for $f, g \in \mathscr{L}^2$, where $\overline{g(x)}$ denotes the complex conjugate of $g(x)$. The exponential or Fourier basis $\{e_j\}_{j\in\mathbb{Z}}$ with $e_j(x) := \exp(2\pi ijx)$ for $x \in [0,1)$ and $j \in \mathbb{Z}$ is an orthonormal basis of $\mathscr{L}^2$. Consequently, any $\xi \in \mathscr{L}^2$ admits an expansion as a discrete Fourier series

$$\xi = \sum_{j\in\mathbb{Z}} \xi_j e_j \tag{4.2.1}$$

with $\xi_j := \langle \xi, e_j\rangle_{\mathscr{L}^2}$ for $j \in \mathbb{Z}$, where the equality (4.2.1) holds in $\mathscr{L}^2$. By Parseval's identity its sequence of Fourier coefficients $\xi_\bullet := (\xi_j)_{j\in\mathbb{Z}}$ is square summable, i.e. it belongs to the Hilbert space $\ell^2 := \ell^2(\mathbb{Z})$ of square-summable complex-valued sequences. The Hilbert space ℓ^2 is equipped with its usual norm $\|\cdot\|_{\ell^2}$ and corresponding inner product $\langle a_\bullet, b_\bullet\rangle_{\ell^2} := \sum_{j\in\mathbb{Z}} a_j\overline{b_j}$ for $a_\bullet, b_\bullet \in \ell^2$. Parseval's identity then states that

$$\|\xi\|_{\mathscr{L}^2} = \|\xi_\bullet\|_{\ell^2}$$

for all $\xi \in \mathscr{L}^2$. For a density function g we further denote by $\mathscr{L}^2(g)$ the set of all real-valued (Borel-measurable) functions h satisfying $\int_0^1 h^2(x)g(x)\mathrm{d}x < \infty$. If $Y \sim g$, then this condition translates to $\mathbb{E}_{Y\sim g}h^2(Y) < \infty$. To be more precise, we define

$$\mathscr{L}^2(g) := \left\{h : [0,1) \to \mathbb{R} : \int_0^1 h^2(x)g(x)\mathrm{d}x < \infty\right\}.$$

We denote the set of square-integrable densities by $\mathcal{D} \in \mathscr{L}^2$ and assume $\varphi, f, g \in \mathcal{D}$.

Fourier coefficients of densities in $\mathcal{D}$. We expand the densities $f, f^\circ \in \mathcal{D} \subseteq \mathscr{L}^2$ in the exponential basis. Since densities $f \in \mathcal{D}$ are normalized to 1, i.e. $1 = \int_0^1 f(x)\mathrm{d}x = \int_0^1 f(x)\overline{\mathbb{1}_{[0,1)}(x)}\mathrm{d}x$, we always have

$$f_0 = 1 \tag{4.2.2}$$

Moreover, since densities are real-valued, we obtain

$$f_j = \overline{f_{-j}} \tag{4.2.3}$$

for all $j \in \mathbb{Z}$. To see this, note that $f(x) = \overline{f(x)}$ for all $x \in \mathbb{R}$. Hence, also $\sum_{k\in\mathbb{Z}} f_k e_k = \sum_{k\in\mathbb{Z}} \overline{f_k}\,\overline{e_k} = \sum_{k\in\mathbb{Z}} \overline{f_k} e_{-k}$. Finally, by projecting onto the j-th basis function and exploiting the orthonormality, we obtain $f_j = \langle \sum_{k\in\mathbb{Z}} f_k e_k, e_j\rangle = \langle \sum_{j\in\mathbb{Z}} \overline{f_k} e_{-k}, e_j\rangle = \overline{f_{-j}}$ for each $j \in \mathbb{Z}$. In particular, using the two properties (4.2.2) and (4.2.3) as well as Parseval's identity, we can rewrite the $\mathscr{L}^2$ distance between f and f° as

$$\|f - f^\circ\|^2_{\mathscr{L}^2} = \|f_\bullet - f^\circ_\bullet\|^2_{\ell^2} = \sum_{j\in\mathbb{Z}} \left|f_j - f^\circ_j\right|^2 = \sum_{|j|\in\mathbb{N}} \left|f_j - f^\circ_j\right|^2 = 2\sum_{j\in\mathbb{N}} \left|f_j - f^\circ_j\right|^2.$$

Moreover, by the circular convolution theorem, the density $g = f \circledast \varphi$ of the observations admits Fourier coefficients satisfying $g_j = f_j \cdot \varphi_j$ for all $j \in \mathbb{N}$. Hence, assuming from here onwards that the Fourier coefficients of the error density are non-vanishing, i.e. $|\varphi_j| > 0$ for all $j \in \mathbb{Z}$, we obtain the representation of the $\mathscr{L}^2$-distance between f and f°

$$\mathrm{q}^2(f - f^\circ) := \sum_{|j|\in\mathbb{N}} \left|f_j - f^\circ_j\right|^2 = \sum_{|j|\in\mathbb{N}} \frac{\left|g_j - g^\circ_j\right|^2}{|\varphi_j|^2},$$

where

$$g_j = \langle g, e_j \rangle_{\mathscr{L}^2} = \int_0^1 g(y)\overline{e_j(y)}\mathrm{d}y = \int_0^1 e_j(-y)g(y)\mathrm{d}y = \mathbb{E}_{Y\sim g}(e_j(-Y))$$

for all $j \in \mathbb{Z}$ and we denote $g^\circ = f^\circ \circledast \varphi$, i.e. $g_j^\circ = f_j^\circ \varphi_j$ for all $j \in \mathbb{Z}$.

Definition of the test statistic. For $k \in \mathbb{N}$ and $[\![k]\!] := [1, k] \cap \mathbb{N}$ let us define an unbiased estimator $\hat{\mathrm{q}}_k^2$ of the truncated version

$$\mathrm{q}_k^2(f - f^\circ) := \sum_{|j| \in [\![k]\!]} \left|f_j - f_j^\circ\right|^2 = \sum_{|j| \in [\![k]\!]} \frac{|g_j|^2}{|\varphi_j|^2} - 2 \sum_{|j| \in [\![k]\!]} \frac{g_j^\circ \overline{g_j}}{|\varphi_j|^2} + \sum_{|j| \in [\![k]\!]} \left|f_j^\circ\right|^2, \tag{4.2.4}$$

where we exploited that due to the symmetry of the summation and the coefficients (4.2.3) we have

$$\sum_{|j| \in [\![k]\!]} \frac{g_j^\circ \overline{g_j}}{|\varphi_j|^2} = \sum_{|j| \in [\![k]\!]} \frac{g_{-j}^\circ \overline{g_{-j}}}{|\varphi_j|^2} = \sum_{|j| \in [\![k]\!]} \frac{\overline{g_j^\circ} g_j}{|\varphi_j|^2}.$$

The first two summands of (4.2.4) are unknown and need to be estimated, the third is known. For the second term, which is a linear term, recall that the Fourier coefficients can be expressed as $g_j = \mathbb{E}_{Y\sim g} e_j(-Y)$. Thus, a natural estimator based on observations $\{Y_l\}_{l=1}^n$ is given by $\frac{1}{n}\sum_{l \in [\![n]\!]} e_j(-Y_l)$. Replacing the unknown Fourier coefficients by their empirical counterparts based on the observations $\{Y_l\}_{l=1}^n$, we obtain

$$\hat{S}_k := \frac{1}{n} \sum_{|j| \in [\![k]\!]} \sum_{l \in [\![n]\!]} \frac{g_j^\circ e_j(Y_l)}{|\varphi_j|^2}$$

as an unbiased estimator of $\sum_{|j| \in [\![k]\!]} \frac{g_j^\circ \overline{g_j}}{|\varphi_j|^2}$. For the first term, which is quadratic, we use the U-statistic

$$\hat{T}_k = \frac{1}{n(n-1)} \sum_{|j| \in [\![k]\!]} \sum_{\substack{l,m \in [\![k]\!] \\ l \neq m}} \frac{e_j(-Y_l) e_j(Y_m)}{|\varphi_j|^2}$$

as an unbiased estimator of $\sum_{|j| \in [\![k]\!]} \frac{|g_j|^2}{|\varphi_j|^2}$. In total, we consider the test statistic

$$\hat{\mathrm{q}}_k^2 := \hat{T}_k - 2\hat{S}_k + \mathrm{q}_k^2(f^\circ).$$

Below, we construct a test that, roughly speaking, compares the estimator to a multiple of its standard deviation.

Decomposition of the test statistic. The key element to analyse the behaviour of the test statistic is the following decomposition

$$\hat{\mathrm{q}}_k^2 = \mathrm{U}_n + 2\mathrm{V}_n + \mathrm{q}_k^2(f - f^\circ) \tag{4.2.5}$$

with the canonical U-statistic

$$\mathrm{U}_n := \frac{1}{n(n-1)} \sum_{|j| \in [\![k]\!]} \sum_{\substack{l,m \in [\![n]\!] \\ l \neq m}} \frac{(e_j(-Y_l) - g_j)(e_j(Y_m) - \overline{g}_j)}{|\varphi_j|^2}, \tag{4.2.6}$$

the centred linear term

$$\mathrm{V}_n := \frac{1}{n} \sum_{|j| \in [\![k]\!]} \sum_{l \in [\![n]\!]} \frac{(g_j - g_j^\circ)(e_j(Y_l) - \overline{g_j})}{|\varphi_j|^2} \tag{4.2.7}$$

and the separation term $\mathrm{q}_k^2(f - f^\circ)$.

Definition of the threshold. In the next proposition we provide bounds for the quantiles of the test statistic $\hat{\mathrm{q}}_k^2$. Define $L_x := (\log(e/x))^{1/2} = (1 - \log(x))^{1/2} \in (1, \infty)$ for $x \in (0, 1)$. Define for $k \in \mathbb{N}$ the quantities

$$\nu_k := \left(\sum_{|j| \in [\![k]\!]} \frac{1}{|\varphi_j|^4} \right)^{1/4} \qquad \text{and} \qquad m_k := \max_{|j| \in [\![k]\!]} |\varphi_j|^{-1}, \tag{4.2.8}$$

which we, roughly speaking, use to characterise the variance of the test statistic $\hat{\mathrm{q}}_k^2$. For $c_1 := 799 \|g_\bullet^\circ\|_{\ell^2} + 1372$, $c_2 := 52 \|g_\bullet^\circ\|_{\ell^1}$ and $\alpha \in (0, 1)$, we define the threshold

$$\tau_k(\alpha) := c_1 \left(1 \vee L_\alpha^2 \sqrt{\frac{\nu_k^2}{n}} \vee L_\alpha^3 \frac{\nu_k^2}{n} \right) L_\alpha \frac{\nu_k^2}{n} + c_2 L_\alpha^2 \frac{m_k^2}{n}. \tag{4.2.9}$$

Note that due to the Cauchy-Schwarz inequality and Parseval's identity, we have $\|g_\bullet^\circ\|_{\ell^1} \leq \|f_\bullet^\circ\|_{\ell^2} \|\varphi_\bullet\|_{\ell^2} = \|f^\circ\|_{\mathscr{L}^2} \|\varphi\|_{\mathscr{L}^2} < \infty$, hence, c_2 is indeed finite.

Proposition 4.2.1 (Bounds for the quantiles of $\hat{\mathrm{q}}_k^2$). For densities $f^\circ, f, \varphi \in \mathcal{D}$ and $n \in \mathbb{N}, n \geq 2$ consider $\{Y_j\}_{j=1}^n \overset{\text{iid}}{\sim} g = f \circledast \varphi$ with joint distribution $\mathbb{P}_f$ and let $g^\circ = f^\circ \circledast \varphi$. Let $\alpha, \beta \in (0, 1)$ and for $k \in \mathbb{N}$ consider the estimator $\hat{\mathrm{q}}_k^2$ and the threshold $\tau_k(\alpha)$ as defined in (4.2.5) and (4.2.9), respectively.

(i) If $\mathscr{L}^2(g^\circ) = \{|\xi|, \xi \in \mathscr{L}^2\}$, then

$$\mathbb{P}_{f^\circ}(\hat{\mathrm{q}}_k^2 \geq \tau_k(\alpha)) \leq \alpha.$$

(ii) If $c_3 := 8 \|g_\bullet\|_{\ell^1} + 826 \|\varphi_\bullet\|_{\ell^2}^2 + 1372$ and the separation condition

$$\mathrm{q}_k^2(f - f^\circ) \geq 2 \left(\tau_k(\alpha) + c_3 L_{\beta/2}^4 \left(1 \vee \frac{\nu_k^2}{n} \right) \frac{\nu_k^2}{n} \right), \tag{4.2.10}$$

holds, then

$$\mathbb{P}_f(\hat{\mathrm{q}}_k^2 < \tau_k(\alpha)) \leq \beta.$$

Proof of Proposition 4.2.1. (i) If $f = f^\circ$ and, hence, $g = g^\circ$, the decomposition (4.2.5) simplifies to $\hat{\mathrm{q}}_k^2 = \mathrm{U}_n$, where U_n is a canonical U-statistic. Applying Proposition D.1.1 of the appendix, a concentration inequality for canonical U-statistics of order 2, with $x = L_\alpha^2 \geq 1$ and quantities $A - D$ satisfying (D.1.2), we obtain

$$\mathbb{P}_{f^\circ} \left(\mathrm{U}_n \geq 8 \frac{C}{n} L_\alpha + 13 \frac{D}{n} L_\alpha^2 + 261 \frac{B}{n^{3/2}} L_\alpha^3 + 343 \frac{A}{n^2} L_\alpha^4 \right) \leq \exp(1 - x). \tag{4.2.11}$$

Consider the quantities $A - C$ defined in (D.2.1) and D in (D.2.2), which under the additional assumption $\mathscr{L}^2(g^\circ) = \{|\xi| : \xi \in \mathscr{L}^2\}$ satisfy (D.1.2) due to Lemma D.2.1. We have

$$
\begin{aligned}
&8\frac{C}{n}L_\alpha + 13\frac{D}{n}L_\alpha^2 + 261\frac{B}{n^{3/2}}L_\alpha^3 + 343\frac{A}{n^2}L_\alpha^4\\
\leq&8\cdot 2\cdot \|g_\bullet\|_{\ell^2}\, L_\alpha \frac{\nu_k^2}{n} + 13\cdot 4\cdot \|g_\bullet\|_{\ell^1}\, L_\alpha^2 \frac{m_k^2}{n} + 261\cdot 3\cdot \|g_\bullet\|_{\ell^2}\, L_\alpha^3 \frac{\nu_k^3}{n^{3/2}} + 343\cdot 4\cdot L_\alpha^4\frac{\nu_k^4}{n^2}\\
=&52\,\|g_\bullet\|_{\ell^1}\, L_\alpha^2\frac{m_k^2}{n} + L_\alpha\frac{\nu_k^2}{n}\left(16\,\|g_\bullet\|_{\ell^2} + 783\,\|g_\bullet\|_{\ell^2}\, L_\alpha^2\frac{\nu_k}{n^{1/2}} + 1372 L_\alpha^3\frac{\nu_k^2}{n}\right)\\
\leq&52\,\|g_\bullet\|_{\ell^1}\, L_\alpha^2\frac{m_k^2}{n} + L_\alpha\frac{\nu_k^2}{n}\left(799\,\|g_\bullet\|_{\ell^2} + 1372\right)\left(1\vee L_\alpha^2\frac{\nu_k}{n^{1/2}}\vee L_\alpha^3\frac{\nu_k^2}{n}\right)\\
=&c_2 L_\alpha^2\frac{m_k^2}{n} + c_1\left(1\vee L_\alpha^2\sqrt{\frac{\nu_k^2}{n}}\vee L_\alpha^3\frac{\nu_k^2}{n}\right)L_\alpha\frac{\nu_k^2}{n}\\
=&\tau_k(\alpha),
\end{aligned}
$$

which together with (4.2.11) shows the assertion (i).

(ii) Keeping the decomposition (4.2.5) in mind, we control the deviations of the U-statistic U_n and the linear statistic V_n by applying Proposition D.1.1 and Lemma D.2.2 of the appendix, respectively. In fact, the quantities $A-D$ given in (D.2.1) of Lemma D.2.1 fulfil (recall that $L_{\beta/2}\geq 1$ for all $\beta>0$)

$$
\begin{aligned}
&8\frac{C}{n}L_{\beta/2} + 13\frac{D}{n}L_{\beta/2}^2 + 261\frac{B}{n^{3/2}}L_{\beta/2}^3 + 343\frac{A}{n^2}L_{\beta/2}^4\\
\leq&8\cdot 2\cdot \|g_\bullet\|_{\ell^2}\, L_{\beta/2}\frac{\nu_k^2}{n} + 13\cdot 2\cdot \|g_\bullet\|_{\ell^2}\, L_{\beta/2}^2\frac{\nu_k^2}{n}\\
&+ 261\cdot 3\cdot \|g_\bullet\|_{\ell^2}\, L_{\beta/2}^3\frac{\nu_k^3}{n^{3/2}} + 343\cdot 4\cdot L_{\beta/2}^4\frac{\nu_k^4}{n^2}\\
\leq&L_{\beta/2}^4\left(42\,\|g_\bullet\|_{\ell^2} + 783\,\|g_\bullet\|_{\ell^2}\frac{\nu_k}{n^{1/2}} + 1372\frac{\nu_k^2}{n}\right)\frac{\nu_k^2}{n}\\
\leq&L_{\beta/2}^4\left(825\,\|g_\bullet\|_{\ell^2} + 1372\right)\left(1\vee\frac{\nu_k}{n^{1/2}}\vee\frac{\nu_k^2}{n}\right)\frac{\nu_k^2}{n}\\
\leq&L_{\beta/2}^4(825\,\|g_\bullet\|_{\ell^2} + 1372)\left(1\vee\frac{\nu_k^2}{n}\right)\frac{\nu_k^2}{n} =: \tau_1,
\end{aligned}
$$

where we exploited that $1\vee a\vee a^2 = 1\vee a^2$ for any $a\geq 0$. Consequently, the event

$$
\Omega_1 := \{\mathrm{U}_n \leq -\tau_1\}
$$

satisfies $\mathbb{P}_f(\Omega_1)\leq \beta/2$ due to Proposition D.1.1 (with the usual symmetry argument). Define further the event

$$
\Omega_2 := \left\{2\mathrm{V}_n \leq -\tau_2 - \frac{1}{2}\mathrm{q}_k^2(f-f^\circ)\right\}
$$

with $\tau_2 := L_{\beta/2}^2(8\,\|g_\bullet\|_{\ell^1} + \|\varphi_\bullet\|_{\ell^2}^2)\left(1\vee\frac{m_k^2}{n}\right)\frac{m_k^2}{n}$. Then we have $\mathbb{P}_f(\Omega_2)\leq\frac{\beta}{2e}\leq\frac{\beta}{2}$ due to Lemma D.2.2 with $x = L_{\beta/2}\geq 1$, which is an application of a Bernstein-type inequality.

We obtain

$$\begin{aligned}
&\tau_1+\tau_2\\
&= L^4_{\beta/2}(825\,\|g_\bullet\|_{\ell^2}+1372)\left(1\vee\frac{\nu_k^2}{n}\right)\frac{\nu_k^2}{n}+L^2_{\beta/2}(8\,\|g_\bullet\|_{\ell^1}+\|\varphi_\bullet\|^2_{\ell^2})\left(1\vee\frac{m_k^2}{n}\right)\frac{m_k^2}{n}\\
&\le L^4_{\beta/2}\left(1\vee\frac{\nu_k^2}{n}\right)\frac{\nu_k^2}{n}\left(825\,\|g_\bullet\|_{\ell^2}+1372+8\,\|g_\bullet\|_{\ell^1}+\|\varphi_\bullet\|^2_{\ell^2}\right)\\
&\le L^4_{\beta/2}c_3\left(1\vee\frac{\nu_k^2}{n}\right)\frac{\nu_k^2}{n}
\end{aligned}$$

with $c_3=8\,\|g_\bullet\|_{\ell^1}+826\,\|\varphi_\bullet\|^2_{\ell^2}+1372$ due to $m_k^2\le\nu_k^2$, $1\le L_{\beta/2}$ and $\|g_\bullet\|_{\ell^2}\le\|\varphi_\bullet\|_{\ell^2}\le\|\varphi_\bullet\|^2_{\ell^2}$. Hence, the assumption (4.2.10) implies

$$\frac{1}{2}\mathrm{q}_k^2(f-f^\circ)\ge\tau_k(\alpha)+\tau_1+\tau_2.$$

The decomposition (4.2.5) yields

$$\begin{aligned}
\mathbb{P}_f\left(\hat{\mathrm{q}}_k^2<\tau_k(\alpha)\right)&=\mathbb{P}_f\left(\left\{\hat{\mathrm{q}}_k^2<\tau_k(\alpha)\right\}\cap\Omega_1\right)+\mathbb{P}_f\left(\left\{\hat{\mathrm{q}}_k^2<\tau_k(\alpha)\right\}\cap\Omega_1^c\right)\\
&\le\mathbb{P}_f(\Omega_1)+\mathbb{P}_f\left(2\mathrm{V}_n+\mathrm{q}_k^2(f-f^\circ)<\tau_k(\alpha)+\tau_1\right)\\
&\le\frac{\beta}{2}+\mathbb{P}_f(\Omega_2)\le\beta,
\end{aligned}$$

which shows (ii) and completes the proof.

□

Remark 4.2.2 (Assumption in (i) of Proposition 4.2.1). *The technical assumption $\mathscr{L}^2(g^\circ)=\{|\xi|:\xi\in\mathscr{L}^2\}$ in Proposition 4.2.1 allows us to express elements of $\mathscr{L}^2(g^\circ)$ in their Fourier expansion. The assumption is needed to obtain the second bound for the quantity D in Lemma D.2.1 of the appendix. It is immediately satisfied for $f^\circ=\mathbb{1}_{[0,1)}$ and if f° is bounded away from 0.* □

Definition of the test. Using the test statistic $\hat{\mathrm{q}}_k^2$ and the threshold $\tau_k(\alpha)$ given in (4.2.5) and (4.2.9), respectively, we define the test

$$\Delta_{k,\alpha}:=\mathbb{1}_{\{\hat{\mathrm{q}}_k^2\ge\tau_k(\alpha)\}},\qquad\text{for } k\in\mathbb{N},\ \alpha\in(0,1).\tag{4.2.12}$$

From (i) in Proposition 4.2.1 it immediately follows that $\Delta_{k,\alpha}$ is a level-α-test for all $k\in\mathbb{N}$. To analyse its power over the alternative, we introduce a regularity constraint, i.e. a nonparametric class of functions $\mathcal{E}=\mathcal{E}^{\mathrm{R}}_{a_\bullet}$, which is formulated in terms of Fourier coefficients. Let $\mathrm{R}>0$ and let $a_\bullet=(a_j)_{j\in\mathbb{N}}$ be a strictly positive, monotonically non-increasing sequence that is bounded by 1. We assume that the differences $f-f^\circ$ belong to the ellipsoid

$$\mathcal{E}^{\mathrm{R}}_{a_\bullet}=\left\{\tilde{f}\in\mathcal{D}:2\sum_{j\in\mathbb{N}}a_j^{-2}\left|\tilde{f}_j\right|^2\le\mathrm{R}^2\right\}.\tag{4.2.13}$$

Note that $\tilde{f}\in\mathcal{E}^{\mathrm{R}}_{a_\bullet}$ imposes conditions on all coefficients $\tilde{f}_j$, $j\in\mathbb{Z}$, since $\left|\tilde{f}_j\right|^2=\left|\tilde{f}_{-j}\right|^2$, $j\in\mathbb{N}$ for all real-valued functions and, additionally, $f_0^\circ=1$ for all densities. The definition (4.2.13) is general enough to cover classes of ordinary and super smooth densities. The second part (ii) of Proposition 4.2.1 now allows to characterise elements in $\mathcal{E}^{\mathrm{R}}_{a_\bullet}$ for which $\Delta_{k,\alpha}$ is powerful. Exploiting these results, we derive an upper bound for the radius of testing of $\Delta_{k,\alpha}$ in terms of ν_k as in (4.2.8) and the regularity parameter $a_\bullet$, that is we define

$$\rho^2_{k,a_\bullet}:=\rho^2_{k,a_\bullet}(n):=a_k^2\vee\frac{\nu_k^2}{n},\tag{4.2.14}$$

where, for now, we suppress the dependence on n in the notation.

Proposition 4.2.3 (Upper bound for the radius of testing of $\Delta_{k,\alpha/2}$). Let $g^\circ = f^\circ \circledast \varphi$ with $f^\circ, \varphi \in \mathscr{L}^2$ satisfy $\mathscr{L}^2(g^\circ) = \{|\xi|, \xi \in \mathscr{L}^2\}$. For $\alpha \in (0,1)$ define

$$\overline{A}_\alpha^2 := \mathrm{R}^2 + 2(8\mathrm{R}\,\|\varphi_\bullet\|_{\ell^2} + 826\,\|\varphi_\bullet\|_{\ell^2}^2 + 859\,\|g_\bullet^\circ\|_{\ell^1} + 2744)L_{\alpha/4}^4. \tag{4.2.15}$$

For all $A \geq \overline{A}_\alpha$ and for all $n, k \in \mathbb{N}$ with $n \geq 2$ and $\nu_k^2 \leq n$, we have

$$\mathcal{R}\left(\Delta_{k,\alpha/2} \mid \mathcal{E}_{a_\bullet}^{\mathrm{R}}, A\rho_{k,a_\bullet}\right) \leq \alpha.$$

Proof of Proposition 4.2.3. We apply Proposition 4.2.1 to show that both the type I and the maximal type II error probability are bounded by $\alpha/2$, then the result follows immediately from the definition of the testing risk

$$\begin{aligned}\mathcal{R}\left(\Delta_{k,\alpha/2} \mid \mathcal{E}_{a_\bullet}^{\mathrm{R}}, A\rho_{k,a_\bullet}\right) &= \mathbb{P}_{f^\circ}\left(\Delta_{k,\alpha/2} = 1\right) + \sup_{f-f^\circ \in \mathscr{L}^2_{A\rho_{k,a_\bullet}} \cap \mathcal{E}_{a_\bullet}^{\mathrm{R}}} \mathbb{P}_f\left(\Delta_{k,\alpha/2} = 0\right) \\ &\leq \alpha/2 + \alpha/2 = \alpha.\end{aligned}$$

Since the assumption of Proposition 4.2.1(i) is fulfilled, the test $\Delta_{k,\alpha/2}$ is a level-$\alpha/2$-test. Hence, for each density $f \in \mathscr{L}^2$ belonging to the alternative, i.e. with $\|f - f^\circ\|_{\mathscr{L}^2}^2 \geq \overline{A}_\alpha^2 \rho_{k,a_\bullet}^2$ and $f - f^\circ \in \mathcal{E}_{a_\bullet}^{\mathrm{R}}$ it remains to verify condition (4.2.10) in order to apply Proposition 4.2.1 (ii) (with $\beta = \alpha/2$). Indeed, in this situations we have

$$\sum_{|j|>k} \left|f_j - f_j^\circ\right|^2 \leq \sum_{|j|>k} \frac{a_k^2}{a_j^2}\left|f_j - f_j^\circ\right|^2 \leq a_k^2 \mathrm{R}^2,$$

since $a_\bullet$ is non-increasing, hence $\frac{a_k}{a_j} \geq 1$ for all $j \geq k$. Thus,

$$\begin{aligned}\mathrm{q}_k^2(f - f^\circ) &= \|f - f^\circ\|_{\mathscr{L}^2}^2 - \sum_{|j|>k} |f - f^\circ|^2 \geq \overline{A}_\alpha^2 \rho_{k,a_\bullet}^2 - a_k^2 \mathrm{R}^2 \\ &\geq 2(8\mathrm{R}\,\|\varphi_\bullet\|_{\ell^2} + 826\,\|\varphi_\bullet\|_{\ell^2}^2 + 859\,\|g_\bullet^\circ\|_{\ell^1} + 2744)L_{\alpha/4}^4 \frac{\nu_k^2}{n}\end{aligned} \tag{4.2.16}$$

where the second inequality is due to the definition of $\overline{A}_\alpha$. Note that using the triangle inequality and the Cauchy-Schwarz inequality, we obtain

$$\|g_\bullet\|_{\ell^1} \leq \|g_\bullet - g_\bullet^\circ\|_{\ell^1} + \|g_\bullet^\circ\|_{\ell^1} \leq \|f_\bullet - f_\bullet^\circ\|_{\ell^2}\,\|\varphi_\bullet\|_{\ell^2} + \|g_\bullet^\circ\|_{\ell^1} \leq \mathrm{R}\,\|\varphi_\bullet\|_{\ell^2} + \|g_\bullet^\circ\|_{\ell^1},$$

where we used

$$\|f_\bullet - f_\bullet^\circ\|_{\ell^2}^2 = \sum_{j \in \mathbb{Z}} \left|f_j - f_j^\circ\right|^2 \leq \sum_{j \in \mathbb{Z}} a_j^{-2}\left|f_j - f_j^\circ\right|^2 \leq \mathrm{R}^2,$$

since $a_\bullet$ is bounded by 1. Hence, (4.2.16) yields

$$\mathrm{q}_k^2(f - f^\circ) \geq 2(8\,\|g_\bullet\|_{\ell^1} + 826\,\|\varphi_\bullet\|_{\ell^2}^2 + 851\,\|g_\bullet^\circ\|_{\ell^1} + 2744)L_{\alpha/4}^4 \frac{\nu_k^2}{n}. \tag{4.2.17}$$

The condition (4.2.10) then follows from (4.2.17) by exploiting further $1 \leq L_{\alpha/2} \leq L_{\alpha/4}$, $\|g_\bullet^\circ\|_{\ell^2} \leq \|g_\bullet^\circ\|_{\ell^1}$ and $m_k^2 \leq \nu_k^2 \leq n$. Indeed,

$$\begin{aligned}\tau_k(\alpha/2) &= c_1\left(1 \vee L_{\alpha/2}^2 \sqrt{\frac{\nu_k^2}{n}} \vee L_{\alpha/2}^3 \frac{\nu_k^2}{n}\right) L_{\alpha/2}\frac{\nu_k^2}{n} + c_2 L_{\alpha/2}^2 \frac{m_k^2}{n} \\ &\leq (c_1 + c_2)L_{\alpha/4}^4 \frac{\nu_k^2}{n}.\end{aligned}$$

Hence,

$$
\begin{aligned}
&2\left(\tau_k(\alpha/2)+c_3L^4_{\alpha/4}\left(1\vee\frac{\nu_k^2}{n}\right)\frac{\nu_k^2}{n}\right)\\
&\quad\le 2(c_1+c_2+c_3)L^4_{\alpha/4}\frac{\nu_k^2}{n}\\
&\quad\le 2(8\,\|g_\bullet\|_{\ell^1}+826\,\|\varphi_\bullet\|^2_{\ell^2}+799\,\|g^\circ_\bullet\|_{\ell^2}+52\,\|g^\circ_\bullet\|_{\ell^1}+2744)L^4_{\alpha/4}\frac{\nu_k^2}{n}\\
&\quad\le 2(8\,\|g_\bullet\|_{\ell^1}+826\,\|\varphi_\bullet\|^2_{\ell^2}+851\,\|g^\circ_\bullet\|_{\ell^1}+2744)L^4_{\alpha/4}\frac{\nu_k^2}{n},
\end{aligned}
$$

which together with (4.2.17) completes the proof. □

Let us introduce a dimension that realizes an optimal bias-variance trade-off and the corresponding radius

$$
k_{a_\bullet}:=k_{a_\bullet}(n):=\arg\min_{k\in\mathbb{N}}\rho^2_{k,a_\bullet}:=\min\left\{k\in\mathbb{N}:\rho^2_{k,a_\bullet}\le\rho^2_{a_\bullet,l}\text{ for all }l\in\mathbb{N}\right\} \tag{4.2.18}
$$

and

$$
\rho^2_{a_\bullet}:=\rho^2_{a_\bullet}(n):=\min_{k\in\mathbb{N}}\rho^2_{k,a_\bullet}=\min_{k\in\mathbb{N}}\left\{a_k^2\vee\frac{\nu_k^2}{n}\right\}.
$$

Corollary 4.2.4 (Upper bound for the radius of testing). Let $g^\circ=f^\circ\circledast\varphi$ with $f^\circ,\varphi\in\mathscr{L}^2$ satisfy $\mathscr{L}^2(g^\circ)=\{|\xi|:\xi\in\mathscr{L}^2\}$. For $\alpha\in(0,1)$ define $\overline{A}^2_\alpha$ as in Proposition 4.2.3. Then for all $A\ge\overline{A}_\alpha$ and $n\ge\sqrt{2}\,|\varphi_1|^{-2}$, we have

$$
\mathcal{R}\left(\Delta_{k_{a_\bullet},\alpha/2}\mid\mathcal{E}^{\mathrm{R}}_{a_\bullet},A\rho_{a_\bullet}\right)\le\alpha. \tag{4.2.19}
$$

Proof of Corollary 4.2.4. The result follows immediately from Proposition 4.2.3, since $\nu^2_{k_{a_\bullet}}\le n$ for all $n\ge\sqrt{2}\,|\varphi_1|^{-2}$. Indeed, $1\ge\rho^2_{a_\bullet,1}\ge\rho^2_{a_\bullet}\ge\nu^2_{k_{a_\bullet}}n^{-1}$ for all $n\ge\sqrt{2}\,|\varphi_1|^{-2}$. □

We shall emphasize that in the case $f^\circ=\mathbb{1}_{[0,1)}$ the radius of testing $\rho_{a_\bullet}$ is known to be minimax (due to the results of Chapter 3), and, hence, the test $\Delta_{k_{a_\bullet},\alpha/2}$ is minimax optimal.

Illustration 4.2.5. We determine the order of the radius of testing

$$
\rho^2_{a_\bullet}(n)=\min_{k\in\mathbb{N}}\left\{a_k^2\vee\frac{\nu_k^2}{n}\right\}
$$

for specific regularity sequences $a_\bullet$ and error densities φ, which are characterised by their sequences of Fourier coefficients $(\varphi_j)_{j\in\mathbb{N}}$ and represent the ill-posedness of the model. For two real-valued sequences $(x_j)_{j\in\mathbb{N}}\in\mathbb{R}^{\mathbb{N}}$ and $(y_j)_{j\in\mathbb{N}}\in\mathbb{R}^{\mathbb{N}}$ we write $x_j\lesssim y_j$ if there exists a constant $c>0$ such that $x_j\le cy_j$ for all $j\in\mathbb{N}$. We write $x_j\sim y_j$, if both $x_j\lesssim y_j$ and $y_j\lesssim x_j$. We distinguish two behaviours of the sequence $a_\bullet$, either polynomial decay $a_j\sim j^{-s}$ for some $s>1/2$, such that $\mathcal{E}^{\mathrm{R}}_{a_\bullet}$ corresponds to a Sobolev ellipsoid of **ordinary smooth** functions, or exponential decay $a_j\sim\exp(-j^s)$ for some $s>0$, where $\mathcal{E}^{\mathrm{R}}_{a_\bullet}$ corresponds to a class of analytic (**super smooth**) functions. The same distinction is made for the regularity of the error density φ. For $p>1/2$ we consider a **mildly ill-posed** model $|\varphi_j|\sim|j|^{-p}$ and for $p>0$ a **severely ill-posed** model $|\varphi_j|\sim\exp(-|j|^p)$. The table below presents the order of the optimal dimension and the upper bound for the radius of testing of the test

$\Delta_{k_{a_\bullet},\alpha/2}$, $\alpha \in (0,1)$ derived in Corollary 4.2.4, which corresponds to the minimax radius of testing . For detailed calculations we refer to Illustration 3.2.6.

Order of the optimal dimension $k_{a_\bullet}$ and the minimax radius $\rho^2_{a_\bullet}$			
a_j (smoothness)	$\lvert\varphi_j\rvert$ (ill-posedness)	$k_{a_\bullet}$	$\rho^2_{a_\bullet}$
j^{-s}	$\lvert j\rvert^{-p}$	$n^{\frac{2}{4p+4s+1}}$	$n^{-\frac{4s}{4s+4p+1}}$
j^{-s}	$e^{-\lvert j\rvert^p}$	$(\log n)^{\frac{1}{p}}$	$(\log n)^{-\frac{2s}{p}}$
e^{-j^s}	$\lvert j\rvert^{-p}$	$(\log n)^{\frac{1}{s}}$	$n^{-1}(\log n)^{\frac{2p+1/2}{s}}$

4.3 Adaptive indirect testing procedure

For a parameter $a_\bullet$ the minimax optimal test $\Delta_{k_{a_\bullet},\alpha/2}$ in Corollary 4.2.4 relies on the dimension parameter $k_{a_\bullet}$, which in turn depends on the smoothness class $\mathcal{E}^{\mathrm{R}}_{a_\bullet}$. Ideally, we want our testing procedure to be *adaptive*, i.e. *assumption-free*, with respect to the alternative class $\mathcal{E}^{\mathrm{R}}_{a_\bullet}$. It should perform optimally for a wide range of alternatives. In this section we therefore propose an adaptive testing procedure by aggregating the tests derived in Section 4.2 over various dimension parameters k. We first generally describe the aggregation procedure and then apply it to the tests defined in (4.2.12).

Description of the adaptation procedure via Bonferroni aggregation. Let $\mathcal{K} \subseteq \mathbb{N}$ be a finite collection of dimension parameters. For $k \in \mathcal{K}$ and levels $(\alpha_k)_{k\in\mathcal{K}} \subseteq (0,1)^{\mathcal{K}}$, we consider the collection of level-α_k-tests $(\phi_{k,\alpha_k})_{k\in\mathcal{K}} = (\mathbb{1}_{\{\zeta_{k,\alpha_k}>0\}})_{k\in\mathcal{K}}$ based on test statistics ζ_{k,α_k}. For $\alpha := \sum_{k\in\mathcal{K}} \alpha_k$, we consider the max-test

$$\phi_{\mathcal{K},\alpha} := \mathbb{1}_{\{\zeta_{\mathcal{K},\alpha}>0\}} \qquad \text{with} \qquad \zeta_{\mathcal{K},\alpha} = \max_{k\in\mathcal{K}} \zeta_{k,\alpha_k},$$

i.e. the test rejects the null hypothesis as soon as one of the tests in the collection does. Under the null hypothesis, we bound the type I error probability of the max-test by the sum of the error probabilities of the individual tests,

$$\mathbb{P}_{f^\circ}(\phi_{\mathcal{K},\alpha} = 1) = \mathbb{P}_{f^\circ}(\max_{k\in\mathcal{K}} \zeta_{k,\alpha_k} > 0) \leq \sum_{k\in\mathcal{K}} \mathbb{P}_{f^\circ}(\zeta_{k,\alpha_k} > 0) \leq \sum_{k\in\mathcal{K}} \alpha_k = \alpha. \tag{4.3.1}$$

Hence, $\phi_{\mathcal{K},\alpha}$ is a level-α-test. Under the alternative, we can bound the type II error probability by the error probability of any of the individual tests,

$$\mathbb{P}_f(\phi_{\mathcal{K},\alpha} = 0) = \mathbb{P}_f(\max_{k\in\mathcal{K}} \zeta_{k,\alpha_k} \leq 0) \leq \min_{k\in\mathcal{K}} \mathbb{P}_f(\zeta_{k,\alpha_k} \leq 0) = \min_{k\in\mathcal{K}} \mathbb{P}_f(\phi_{k,\alpha_k} = 0). \tag{4.3.2}$$

Therefore, $\phi_{\mathcal{K},\alpha}$ has the maximal power achievable by a test in the collection. The bounds (4.3.1) and (4.3.2) have opposing effects on the choice of the collection $\mathcal{K}$. On the one hand, it should be as small as possible to keep the type I error probability small. On the other hand, it must be large enough to contain an optimal dimension parameter $k_{a_\bullet}$ for a wide range of smoothness parameters $a_\bullet$. Typically, there is a cost to pay for adaptation, which is characterized by the size of a *(minimal) adaptive factor*, defined in Section 4.1. Let us heuristically explain what causes the adaptive factor and give a reason for its typical order.

Typical order of the adaptive factor. We have seen that the minimax optimal test introduced in Section 4.2, roughly speaking, compares an estimator $\hat{\mathrm{q}}_k^2$ to a multiple (depending on α) of its standard deviation, which is typically of order $\frac{\nu_k^2}{n}$, i.e. we reject the null as soon as $\hat{\mathrm{q}}_k^2 \gtrsim C_\alpha \frac{\nu_k^2}{n}$. If the deviations of the estimator follow a (sub)Gaussian regime, the threshold constant C_α can be chosen to be of order $\sqrt{\log(1/\alpha)}$ to guarantee level-α. We consider the Bonferroni correction of these tests, that is, the error levels $\alpha_k = \frac{\alpha}{|\mathcal{K}|}$. A natural choice for the collection of dimension parameters $\mathcal{K}$ is a geometric grid $\mathcal{K} = \left\{2^0, 2^1, \dots, 2^{\lfloor 2\log_2 n\rfloor}\right\}$, since in many cases $k_{a_\bullet} \leq n^2$, which yields $|\mathcal{K}| \sim \log n$. Then, the *new* thresholds behave like

$$C_{\alpha_k}\frac{\nu_k^2}{n} \sim \tilde{C}_\alpha \frac{\sqrt{\log\log n}}{n}\nu_k^2,$$

i.e. the *effective sample size* is reduced to $\delta_n n = \frac{n}{\sqrt{\log\log n}}$ by the adaptive factor $\delta_n^{-1} = \sqrt{\log\log n}$.

4.3.1 Aggregation of the indirect tests and the choice of the levels α_k

Denote by $\mathcal{A} \subseteq \mathbb{R}^{\mathbb{N}}_{>0}$ a set of strictly positive, monotonically non-increasing sequences bounded by 1. The set $\mathcal{A}$ characterises the collection of alternatives $\left\{\mathcal{E}^{\mathrm{R}}_{a_\bullet} : a_\bullet \in \mathcal{A}\right\}$, for which we analyse the power of our testing procedure simultaneously. Let $\mathcal{K} \subseteq \mathbb{N}$ with $|\mathcal{K}| < \infty$ and $\alpha \in (0,1)$. We apply the aggregation described above and obtain a max-test with a Bonferroni choice of error levels

$$\Delta_{\mathcal{K},\alpha} := \mathbb{1}_{\{\hat{Q}_{\mathcal{K},\alpha}>0\}} \qquad \text{with} \qquad \hat{Q}_{\mathcal{K},\alpha} := \max_{k\in\mathcal{K}}\left(\hat{\mathrm{q}}_k^2 - \tau_k(\tfrac{\alpha}{|\mathcal{K}|})\right),$$

where $\hat{\mathrm{q}}_k^2$ and $\tau_k(\alpha)$ are the test statistic and the threshold of the indirect test defined in (4.2.5) and (4.2.9), respectively. In this paper we consider a classical Bonferroni choice of error levels, $\alpha_k = \frac{\alpha}{|\mathcal{K}|}$. In the next remark we discuss other possible aggregation choices and compare them to our method. The Monte-Carlo quantile and threshold method is e.g. used in Laurent et al. [2003] and Fromont and Laurent [2006].

Remark 4.3.1 (Choice of $(\alpha_k)_{k\in\mathcal{K}}$). *Let us describe three different methods for choosing thresholds for the statistics $\hat{\mathrm{q}}_k^2$ and the levels $(\alpha_k)_{k\in\mathcal{K}}$.*
Monte-Carlo quantile method. *Roughly speaking, instead of using the thresholds $\tau_k(\alpha)$ that we introduce in (4.2.9), this approach uses the (unspecified) quantiles of $\hat{\mathrm{q}}_k^2$ under the null hypothesis. Let us be more precise and denote by $t_k(\alpha)$ the $(1-\alpha)$-quantile of $\hat{\mathrm{q}}_k^2$ under the null hypothesis $f = f^\circ$. Let*

$$\alpha^\star := \sup\left\{u \in (0,1) : \mathbb{P}_{f^\circ}\left(\max_{k\in\mathcal{K}}\left(\hat{\mathrm{q}}_k^2 - t_k(u)\right) > 0\right) \leq \alpha\right\}$$

and consider the test statistic and the corresponding test

$$\Delta^\star_{\mathcal{K},\alpha} := \mathbb{1}_{\{T^\star_{\mathcal{K},\alpha^\star}>0\}} \qquad \textit{with} \qquad T^\star_{\mathcal{K},\alpha^\star} := \max_{k\in\mathcal{K}}\left(\hat{\mathrm{q}}_k^2 - t_k(\alpha^\star)\right)$$

Then, by definition

$$\mathbb{P}_{f^\circ}\left(\Delta^\star_{\mathcal{K},\alpha} = 1\right) = \mathbb{P}_{f^\circ}\left(\max_{k\in\mathcal{K}}\left(\hat{\mathrm{q}}_k^2 - t_k(\alpha^\star)\right) > 0\right) \leq \alpha.$$

The drawback of this method is that in general there are no explicit formulas for the quantiles $t_k(u), k \in \mathcal{K},\ u \in (0,1)$ and the chosen error level $\alpha^\star$. Therefore, in practice they have to be determined e.g. via a Monte-Carlo-simulation.

Monte-Carlo threshold method. *Instead of using the quantiles* $t_k(u)$*, we use the explicit upper bounds for the quantiles* $\tau_k(u) \geq t_k(u)$ *for* $u \in (0,1)$ *that we determined in Proposition 4.2.1. Let*

$$\alpha^\dagger := \sup\left\{u \in (0,1) : \mathbb{P}_{f^\circ}\left(\max_{k\in\mathcal{K}}\left(\hat{\mathrm{q}}_k^2 - \tau_k(u)\right) > 0\right) \leq \alpha\right\}$$

and consider the test statistic and the corresponding test

$$\Delta^\dagger_{\mathcal{K},\alpha} := \mathbb{1}_{\left\{T^\dagger_{\mathcal{K},\alpha^\dagger} > 0\right\}} \qquad \text{with} \qquad T^\dagger_{\mathcal{K},\alpha^\dagger} := \max_{k\in\mathcal{K}}\left(\hat{\mathrm{q}}_k^2 - \tau_k(\alpha^\dagger)\right).$$

Then, by definition

$$\mathbb{P}_{f^\circ}\left(\Delta^\dagger_{\mathcal{K},\alpha} = 1\right) = \mathbb{P}_{f^\circ}\left(\max_{k\in\mathcal{K}}\left(\hat{\mathrm{q}}_k^2 - \tau_k(\alpha^\dagger)\right) > 0\right) \leq \alpha.$$

This method no longer requires simulations for the thresholds. However, there is still no explicit formula for $\alpha^\dagger$*, which again has to be determined via a Monte-Carlo-simulation.*
Bonferroni method. *We simply define*

$$\alpha_k := \frac{\alpha}{|\mathcal{K}|}$$

and consider the test statistic and the corresponding test

$$\Delta_{\mathcal{K},\alpha} := \mathbb{1}_{\{\hat{T}_{\mathcal{K},\alpha} > 0\}} \qquad \text{with} \qquad \hat{T}_{\mathcal{K},\alpha} = \max_{k\in\mathcal{K}}\left(\hat{T}_k - \tau_k(\alpha_k)\right).$$

Then, by definition

$$\mathbb{P}_{f^\circ}\left(\Delta_{\mathcal{K},\alpha} = 1\right) = \mathbb{P}_{f^\circ}\left(\max_{k\in\mathcal{K}}\left(\hat{\mathrm{q}}_k^2 - \tau_k(\alpha_k)\right) > 0\right) \leq \sum_{k\in\mathcal{K}} \alpha_k = \sum_{k\in\mathcal{K}} \frac{\alpha}{|\mathcal{K}|} = \alpha.$$

While it is a more conservative method, it allows us to explicitly show the dependence of the testing radius of the max*-test on the size of the collection* $\mathcal{K}$*. This dependence is naturally also present in the Monte Carlo methods, though hidden in the definition of* $\alpha^\star$ *resp.* $\alpha^\dagger$*.*
The power of the three methods. *All three methods yield* max*-tests of level-*α*. To analyse their power, we note that for any* $f \in \mathcal{D}$

$$\begin{aligned}
\mathbb{P}_f\left(\Delta^\star_{\alpha,k} = 0\right) &\leq \min_{k\in\mathcal{K}} \mathbb{P}_f\left(\hat{\mathrm{q}}_k^2 < t_k(\alpha^\star)\right),\\
\mathbb{P}_f\left(\Delta^\dagger_{\alpha,k} = 0\right) &\leq \min_{k\in\mathcal{K}} \mathbb{P}_f\left(\hat{\mathrm{q}}_k^2 < \tau_k(\alpha^\dagger)\right),\\
\mathbb{P}_f\left(\Delta_{\alpha,k} = 0\right) &\leq \min_{k\in\mathcal{K}} \mathbb{P}_f\left(\hat{\mathrm{q}}_k^2 < \tau_k(\alpha_k)\right).
\end{aligned} \tag{4.3.3}$$

In particular it follows from the definition of $\alpha^\star$*,* $\alpha^\dagger$ *and the fact that* $\tau_k(\alpha)$ *is monotonically increasing in* α *that*

$$t_k(\alpha^\star) \leq \tau_k(\alpha^\dagger) \leq \tau_k(\alpha_k)$$

*Hence, showing via (4.3.3) that the Bonferroni-*max*-test* $\Delta_{\alpha,k}$ *is powerful for an element* $f \in \mathcal{D}$ *immediately implies that the other two* max*-tests are also powerful.* □

4.3.2 Testing radius of the indirect max-test

The next proposition determines an adaptive upper bound for the radius of testing of the max-test. The upper bound essentially has two regimes. The adaptive factor δ^{-1} depends on which regime determines the behaviour of the radius of testing. For the Bonferroni choice of error levels, $\alpha_k = \frac{\alpha}{|\mathcal{K}|}$, the adaptive factor δ^{-1} is in all cases of order $(\log|\mathcal{K}|)^c$ for $c \in \left\{\frac{1}{2}, 1\right\}$. Below, we give conditions for which the adaptive factor is of order $(\log|\mathcal{K}|)^{1/2}$, which we show to be the minimal adaptive factor.

Recall that the max-test $\Delta_{\mathcal{K},\alpha}$ only aggregates over a finite set $\mathcal{K} \subseteq \mathbb{N}$. We define the minimal achievable radius of testing over the set $\mathcal{K}$ as

$$\rho^2_{\mathcal{K},a_\bullet}(n) := \min_{k\in\mathcal{K}} \rho^2_{k,a_\bullet}(n) \qquad \text{with} \qquad \rho^2_{k,a_\bullet}(n) := a_k^2 \vee \frac{\nu_k^2}{n},$$

with ν_k^2 as in (4.2.8) and a regularity parameter $a_\bullet = (a_j)_{j\in\mathbb{N}} \in \mathcal{A}$. Since $\rho^2_{a_\bullet}(n)$ in (4.2.14) is defined as the minimum taken over $\mathbb{N}$ instead of $\mathcal{K}$, for $n \in \mathbb{N}$ we always have

$$\rho^2_{a_\bullet}(n) = \rho^2_{\mathbb{N},a_\bullet}(n) \le \rho^2_{\mathcal{K},a_\bullet}(n).$$

Moreover, replacing ν_k^2 by m_k^2 as in (4.2.8), let us define a remainder radius, typically negligible compared to $\rho^2_{\mathcal{K},a_\bullet}(n)$,

$$r^2_{\mathcal{K},a_\bullet}(n) := \min_{k\in\mathcal{K}} r^2_{k,a_\bullet}(n) \qquad \text{with} \qquad r^2_{k,a_\bullet}(n) := a_k^2 \vee \frac{m_k^2}{n}. \tag{4.3.4}$$

Proposition 4.3.2 (Uniform radius of testing over $\mathcal{A}$).
Under the assumptions of Proposition 4.2.1, let $\alpha \in (0,1)$ and consider $\overline{A}_\alpha$ as in (4.2.15). Then, for all $A \ge \overline{A}_\alpha$ and for all $n \in \mathbb{N}$, $n \ge 2$,

$$\sup_{a_\bullet\in\mathcal{A}} \mathcal{R}\left(\Delta_{\mathcal{K},\alpha/2} \mid \mathcal{E}^{\mathrm{R}}_{a_\bullet}, A\left(1 \vee \frac{\rho_{\mathcal{K},a_\bullet}(\delta n)}{\delta^{3/2}}\right)\left(r_{\mathcal{K},a_\bullet}(\delta^2 n) \vee \rho_{\mathcal{K},a_\bullet}(\delta n)\right)\right) \le \alpha$$

with $\delta = (1 \vee \log|\mathcal{K}|)^{-1/2}$.

Proof of Proposition 4.3.2. For each $a_\bullet \in \mathcal{A}$ we apply Proposition 4.2.1 to show that both the type I and the maximal type II error probability are bounded by $\alpha/2$. The result then follows immediately from the definition of the testing risk

$$\begin{aligned}\mathcal{R}\left(\Delta_{\mathcal{K},\alpha/2} \mid \mathcal{E}^{\mathrm{R}}_{a_\bullet}, A\rho\right) &= \mathbb{P}_{f^\circ}\left(\Delta_{\mathcal{K},\alpha/2} = 1\right) + \sup_{f-f^\circ\in\mathscr{L}^2_{A\rho}\cap\mathcal{E}^{\mathrm{R}}_{a_\bullet}} \mathbb{P}_f\left(\Delta_{\mathcal{K},\alpha/2} = 0\right)\\ &\le \alpha/2 + \alpha/2 = \alpha.\end{aligned}$$

Under the null hypothesis, the claim follows from (4.3.1) together with Proposition 4.2.1 (i) and $\sum_{k\in\mathcal{K}} \alpha_k = \sum_{k\in\mathcal{K}} \frac{\alpha}{2|\mathcal{K}|} = \frac{\alpha}{2}$. Under the alternative, let $f \in \mathscr{L}^2$ with $f - f^\circ \in \mathcal{E}^{\mathrm{R}}_{a_\bullet}$ satisfy

$$\|f - f^\circ\|^2_{\mathscr{L}^2} \ge \overline{A}^2_\alpha \left(1 \vee \frac{\rho^2_{\mathcal{K},a_\bullet}(\delta n)}{\delta^3}\right)\left(r^2_{\mathcal{K},a_\bullet}(\delta^2 n) \vee \rho^2_{\mathcal{K},a_\bullet}(\delta n)\right). \tag{4.3.5}$$

It is sufficient to use the elementary bound (4.3.2) together with the following two observations, which we show below:

1. Whenever $f \in \mathscr{L}^2$ satisfies $f - f^\circ \in \mathcal{E}^{\mathrm{R}}_{a_\bullet}$ and

$$\|f - f^\circ\|^2_{\mathscr{L}^2} \geq \overline{A}^2_\alpha \left(a_k^2 \vee \left(1 \vee \frac{\nu_k}{\delta^2 n^{1/2}} \vee \frac{\nu_k^2}{\delta^3 n}\right) \frac{\nu_k^2}{\delta n} \vee \frac{m_k^2}{\delta^2 n} \right), \tag{4.3.6}$$

then

$$\mathbb{P}_f \left(\hat{\mathrm{q}}_k^2 < \tau_k \left(\tfrac{\alpha}{2|\mathcal{K}|} \right) \right) \leq \frac{\alpha}{2}.$$

2. If the separation condition (4.3.5) is satisfied, then there exists a $k \in \mathcal{K}$ such that (4.3.6) is fulfilled.

Consequently, we have

$$\mathbb{P}_f(\Delta_{\mathcal{K},\alpha/2} = 0) \leq \min_{k \in \mathcal{K}} \mathbb{P}_f(\Delta_{k,\alpha_k/2} = 0) \leq \frac{\alpha}{2}.$$

for all $f \in \mathscr{L}^2$ satisfying $f - f^\circ \in \mathcal{E}^{\mathrm{R}}_{a_\bullet}$ and (4.3.5). Thus, the maximal type II error probability is also bounded by $\alpha/2$. It remains to show (1.) and (2.).

1. The claim follows from Proposition 4.2.1 (ii) (with $\beta = \alpha/2$), since (4.3.6) implies (4.2.10), which states

$$\mathrm{q}_k^2(f - f^\circ) \geq 2 \left(\tau_k(\tfrac{\alpha}{2|\mathcal{K}|}) + c_3 L^4_{\alpha/4} \left(1 \vee \frac{\nu_k^2}{n}\right) \frac{\nu_k^2}{n} \right), \tag{4.3.7}$$

with $\tau_k(\frac{\alpha}{2|\mathcal{K}|})$ as in (4.2.9). Indeed, exploiting $L^2_{\alpha/2} = \log(2e/\alpha) \geq 1$ and, hence,

$$L^2_{\alpha/(2|\mathcal{K}|)} = \log(2e\,|\mathcal{K}|\,/\alpha) = \log|\mathcal{K}| + L^2_{\alpha/2} \leq L^2_{\alpha/2}(1 + \log|\mathcal{K}|) = L^2_{\alpha/2}\delta^{-2},$$

we have

$$\begin{aligned}
\tau_k(\tfrac{\alpha}{2|\mathcal{K}|}) &= c_1 \left(1 \vee L^2_{\alpha/(2|\mathcal{K}|)} \sqrt{\frac{\nu_k^2}{n}} \vee L^3_{\alpha/(2|\mathcal{K}|)} \frac{\nu_k^2}{n}\right) L_{\alpha/(2|\mathcal{K}|)} \frac{\nu_k^2}{n} + c_2 L^2_{\alpha/(2|\mathcal{K}|)} \frac{m_k^2}{n} \\
&\leq c_1 \left(1 \vee L^2_{\alpha/2} \sqrt{\frac{\nu_k^2}{\delta^4 n}} \vee L^3_{\alpha/2} \frac{\nu_k^2}{\delta^3 n}\right) L_{\alpha/2} \frac{\nu_k^2}{\delta n} + c_2 L^2_{\alpha/2} \frac{m_k^2}{\delta^2 n} \\
&\leq c_1 L^4_{\alpha/2} \left(1 \vee \sqrt{\frac{\nu_k^2}{\delta^4 n}} \vee \frac{\nu_k^2}{\delta^3 n}\right) \frac{\nu_k^2}{\delta n} + c_2 L^2_{\alpha/2} \frac{m_k^2}{\delta^2 n} \\
&\leq (c_1 + c_2) L^4_{\alpha/2} \left(\left(1 \vee \sqrt{\frac{\nu_k^2}{\delta^4 n}} \vee \frac{\nu_k^2}{\delta^3 n}\right) \frac{\nu_k^2}{\delta n} \vee \frac{m_k^2}{\delta^2 n} \right).
\end{aligned}$$

Additionally using $L_{\alpha/4} \geq L_{\alpha/2}$, $1 \geq \delta$ the right-hand side of (4.3.7) is bounded by

$$\begin{aligned}
&2 \left(\tau_k(\tfrac{\alpha}{2|\mathcal{K}|}) + c_3 L^4_{\alpha/4} \left(1 \vee \frac{\nu_k^2}{n}\right) \frac{\nu_k^2}{n} \right) \\
&\leq 2(c_1 + c_2) L^4_{\alpha/2} \left(\left(1 \vee \sqrt{\frac{\nu_k^2}{\delta^4 n}} \vee \frac{\nu_k^2}{\delta^3 n}\right) \frac{\nu_k^2}{\delta n} \vee \frac{m_k^2}{\delta^2 n} \right) + 2c_3 L^4_{\alpha/4} \left(1 \vee \frac{\nu_k^2}{n}\right) \frac{\nu_k^2}{n} \\
&\leq 2(c_1 + c_2 + c_3) L^4_{\alpha/4} \left(\left(1 \vee \sqrt{\frac{\nu_k^2}{\delta^4 n}} \vee \frac{\nu_k^2}{\delta^3 n}\right) \frac{\nu_k^2}{\delta n} \vee \frac{m_k^2}{\delta^2 n} \right).
\end{aligned}$$

Hence, since $\overline{A}_\alpha^2 - R^2 \geq 2(c_1 + c_2 + c_3)L_{\alpha/4}^4$ the condition (4.2.10) of Proposition 4.2.1 holds whenever

$$\mathrm{q}_k^2(f - f^\circ) \geq \left(\overline{A}_\alpha^2 - R^2\right)\left(\left(1 \vee \sqrt{\frac{\nu_k^2}{\delta^4 n}} \vee \frac{\nu_k^2}{\delta^3 n}\right)\frac{\nu_k^2}{\delta n} \vee \frac{m_k^2}{\delta^2 n}\right). \tag{4.3.8}$$

Let us verify this condition. Due to $f - f^\circ \in \mathcal{E}_{a_\bullet}^{\mathrm{R}}$ and, hence, $\sum_{|j|>k}\left|f_j^\circ - f_j\right|^2 \leq a_k^2 \mathrm{R}^2$, the condition (4.3.6) implies

$$\begin{aligned}\mathrm{q}_k^2(f - f^\circ) &= \|f - f^\circ\|_{\mathscr{L}^2}^2 - \sum_{|j|>k}\left|f_j^\circ - f_j\right|^2 \\ &\geq \left(\overline{A}_\alpha^2 - R^2\right)\left(\left(1 \vee \sqrt{\frac{\nu_k^2}{\delta^4 n}} \vee \frac{\nu_k^2}{\delta^3 n}\right)\frac{\nu_k^2}{\delta n} \vee \frac{m_k^2}{\delta^2 n}\right),\end{aligned}$$

which justifies the application of Proposition 4.2.1. If (4.3.6) is satisfied, then also (4.3.8) and thus (4.2.10), which shows the claim (1.).

2. By the Balancing Lemma A.2.1 we have

$$r_{\mathcal{K},a_\bullet}^2(\delta^2 n) \vee \rho_{\mathcal{K},a_\bullet}^2(\delta n) = a_k^2 \vee \frac{m_k^2}{\delta^2 n} \vee \frac{\nu_k^2}{\delta n} \quad \text{and} \quad \rho_{\mathcal{K},a_\bullet}^2(\delta n) \geq \frac{\nu_k^2}{\delta n}$$

for at least one $k \in \mathcal{K}$. Hence, there exists a dimension parameter $k \in \mathcal{K}$ such that

$$\left(r_{\mathcal{K},a_\bullet}^2(\delta^2 n) \vee \rho_{\mathcal{K},a_\bullet}^2(\delta n)\right)\left(1 \vee \frac{\rho_{\mathcal{K},a_\bullet}^2(\delta n)}{\delta^3}\right) \geq a_k^2 \vee \frac{m_k^2}{\delta^2 n} \vee \left(\frac{\nu_k^2}{\delta n}\left(1 \vee \frac{\nu_k^2}{\delta^4 n}\right)\right).$$

Since

$$1 \vee \frac{\nu_k^2}{\delta^4 n} \geq 1 \vee \frac{\nu_k}{\delta^2 n^{1/2}} \vee \frac{\nu_k^2}{\delta^3 n},$$

this shows (4.3.6) and, hence, (2.), which completes the proof.

□

Corollary 4.3.3 (Worst-case adaptive factor). Under the assumptions of Proposition 4.2.1, let $\alpha \in (0, 1)$ and consider $\overline{A}_\alpha$ as in (4.2.15). Then, for all $A \geq \overline{A}_\alpha$ and for all $n \in \mathbb{N}$, $n \geq 2$,

$$\sup_{a_\bullet \in \mathcal{A}} \mathcal{R}\left(\Delta_{\mathcal{K},\alpha/2} \mid \mathcal{E}_{a_\bullet}^{\mathrm{R}}, A\left(1 \vee \rho_{\mathcal{K},a_\bullet}(\delta^2 n)\right)\rho_{\mathcal{K},a_\bullet}(\delta^2 n)\right) \leq \alpha$$

with $\delta = (1 \vee \log|\mathcal{K}|)^{-1/2}$.

Proof of Corollary 4.3.3. The proof follows along the lines of the proof of Proposition 4.3.2, considering

$$\|f - f^\circ\|_{\mathscr{L}^2}^2 \geq \overline{A}_\alpha^2\left(1 \vee \rho_{\mathcal{K},a_\bullet}^2(\delta^2 n)\right)\rho_{\mathcal{K},a_\bullet}^2(\delta^2 n) \tag{4.3.9}$$

instead of (4.3.5). In fact, as in the proof of Proposition 4.3.2, it is sufficient to show (2.) under the separation condition (4.3.9). For each $a_\bullet \in \mathcal{A}$ under (4.3.9) the dimension parameter $k_\star := \arg\min_{k\in\mathcal{K}} \rho^2_{k,a_\bullet}(\delta^2 n)$ satisfies

$$\begin{aligned}\left(1 \vee \rho^2_{\mathcal{K},a_\bullet}(\delta^2 n)\right) \rho^2_{\mathcal{K},a_\bullet}(\delta^2 n) &= \left(1 \vee \rho_{\mathcal{K},a_\bullet}(\delta^2 n) \vee \rho^2_{\mathcal{K},a_\bullet}(\delta^2 n)\right) \rho^2_{\mathcal{K},a_\bullet}(\delta^2 n)\\ &\geq a^2_{k_\star} \vee \frac{\nu^2_{k_\star}}{\delta^2 n}\left(1 \vee \frac{\nu_{k_\star}}{\delta n^{1/2}} \vee \frac{\nu^2_{k_\star}}{\delta^2 n}\right)\\ &\geq a^2_{k_\star} \vee \frac{m^2_{k_\star}}{\delta^2 n} \vee \frac{\nu^2_{k_\star}}{\delta n}\left(\frac{1}{\delta} \vee \frac{\nu_{k_\star}}{\delta^2 n^{1/2}} \vee \frac{\nu^2_{k_\star}}{\delta^3 n}\right)\\ &\geq a^2_{k_\star} \vee \frac{m^2_{k_\star}}{\delta^2 n} \vee \frac{\nu^2_{k_\star}}{\delta n}\left(1 \vee \frac{\nu_{k_\star}}{\delta^2 n^{1/2}} \vee \frac{\nu^2_{k_\star}}{\delta^3 n}\right)\end{aligned}$$

since $a^2_{k_\star} \vee \frac{\nu^2_{k_\star}}{\delta^2 n} = \rho^2_{\mathcal{K},a_\bullet}(\delta^2 n)$, $\delta \leq 1$ and $m^2_{k_\star} \leq \nu^2_{k_\star}$. This shows (4.3.6) and, consequently, (2.). We obtain the assertion by proceeding exactly as in the proof of Proposition 4.3.2. □

By Corollary 4.3.3, $\rho^2_{\mathcal{K},a_\bullet}(\delta^2 n)$ is an upper bound for the radius of testing of the indirect max-test as soon as $\rho^2_{\mathcal{K},a_\bullet}(\delta^2 n) \leq 1$. The latter is satisfied for an arbitrary regularity parameter $a_\bullet \in \mathcal{A}$, if $1 \in \mathcal{K}$ (condition on the class $\mathcal{K}$) and $n \geq \sqrt{2}\,|\varphi_1|^{-2}\,(1 + \log|\mathcal{K}|)$ (condition on the minimal sample size). Indeed, under the two conditions, we have

$$\rho^2_{\mathcal{K},a_\bullet}(\delta^2 n) = \min_{k\in\mathcal{K}}\left\{a^2_k \vee \frac{\nu^2_k}{\delta^2 n}\right\} \leq a^2_1 \vee \frac{\nu^2_1}{\delta^2 n} \leq 1 \vee \frac{\sqrt{2}\,|\varphi_1|^{-2}}{\delta^2 n} = 1.$$

Hence, in this case we obtain an **adaptive factor of order** $\log|\mathcal{K}|$**.** The next corollary establishes $\rho^2_{\mathcal{K},a_\bullet}(\delta n)$ as a sharper upper bound for the radius of testing of the indirect max-test under additional conditions, all of which are e.g. satisfied in the examples considered below in Illustration 4.3.6. Therefore, under these additional conditions we obtain an **adaptive factor of order** $\sqrt{\log|\mathcal{K}|}$**.**

Corollary 4.3.4 (Best-case adaptive factor). Under the assumptions of Proposition 4.2.1, let $\alpha \in (0,1)$ and consider $\overline{A}_\alpha$ as in (4.2.15). If there exist constants $c, C > 1$ such that

$$r_{\mathcal{K},a_\bullet}(\delta^2 n) \leq c\rho_{\mathcal{K},a_\bullet}(\delta n) \qquad \text{and} \qquad \rho_{\mathcal{K},a_\bullet}(\delta n) \leq C\delta^{3/2} \tag{4.3.10}$$

for all $a_\bullet \in \mathcal{A}$, then for all $A \geq c \cdot C \cdot \overline{A}_\alpha$ and for all $n \in \mathbb{N}$, $n \geq 2$,

$$\sup_{a_\bullet\in\mathcal{A}} \mathcal{R}\left(\Delta_{\mathcal{K},\alpha/2} \mid \mathcal{E}^{\mathrm{R}}_{a_\bullet}, A\rho_{\mathcal{K},a_\bullet}(\delta n)\right) \leq \alpha$$

with $\delta = (1 \vee \log|\mathcal{K}|)^{-1/2}$.

Proof of Corollary 4.3.4. Under the assumptions (4.3.10) we have

$$\begin{aligned}\left(1 \vee \frac{\rho_{\mathcal{K},a_\bullet}(\delta n)}{\delta^{3/2}}\right)\left(r_{\mathcal{K},a_\bullet}(\delta^2 n) \vee \rho_{\mathcal{K},a_\bullet}(\delta n)\right) &\leq \left(1 \vee \frac{\rho_{\mathcal{K},a_\bullet}(\delta n)}{\delta^{3/2}}\right)(c\rho_{\mathcal{K},a_\bullet}(\delta n) \vee \rho_{\mathcal{K},a_\bullet}(\delta n))\\ &\leq (1 \vee C)\,(c\rho_{\mathcal{K},a_\bullet}(\delta n) \vee \rho_{\mathcal{K},a_\bullet}(\delta n))\\ &\leq c \cdot C \cdot \rho_{\mathcal{K},a_\bullet}(\delta n),\end{aligned}$$

hence, the assertion follows directly from Proposition 4.3.2. □

Remark 4.3.5 (Choice of the collection $\mathcal{K}$). *Ideally, the collection $\mathcal{K} \subseteq \mathbb{N}$ is chosen such that its elements approximate the optimal parameter $k_{a_\bullet}$ for all $a_\bullet \in \mathcal{A}$ given in (4.2.18) sufficiently well. Note that $k_{a_\bullet} \leq \frac{n^2}{2}$ for n reasonably large (precisely $n \geq \sqrt{2}\,|\varphi_1|^{-2}$, which implies $1 \geq a_1^2 \vee \frac{\nu_k^2}{n} = \rho_{1,a_\bullet}^2 \geq \rho_{a_\bullet}^2 \geq \frac{\nu_{k_{a_\bullet}}^2}{n} \geq \frac{\sqrt{2k_{a_\bullet}}}{n}$), Hence, a naive choice is*

$$\mathcal{K}_1 = \left\{1, \dots, \lfloor \tfrac{n^2}{2} \rfloor\right\} \qquad \text{with} \qquad |\mathcal{K}_1| = \lfloor \tfrac{n^2}{2} \rfloor,$$

which yields an adaptive factor of order $(\log n)^{1/2}$. However, in most cases, a minimisation over a geometric grid

$$\mathcal{K}_2 = \left\{2^j, j \in \left\{0, \dots, \lfloor \log_2(\tfrac{n^2}{2}) \rfloor\right\}\right\} \qquad \text{with} \qquad |\mathcal{K}_2| = \lfloor \log_2(\tfrac{n^2}{2}) \rfloor$$

approximates the minimisation over $\mathbb{N}$ well enough. The resulting adaptive factor is then of order $(\log\log n)^{1/2}$. For some special cases the even smaller collection

$$\mathcal{K}_s = \left\{2^j, j \in \left\{0, \dots, \lfloor \tfrac{\log_2 \log n}{s} \rfloor\right\}\right\} \qquad \text{with} \qquad |\mathcal{K}_s| = \lfloor \tfrac{\log_2 \log n}{s} \rfloor$$

(compare Illustration 4.3.6 below) is still sufficient, resulting in an adaptive factor of order $(\log\log\log n)^{1/2}$. □

Illustration 4.3.6. For the typical configurations for regularity and ill-posedness introduced in Illustration 4.2.5, the tables below display the upper bounds for the adaptive radii of the max-tests $\Delta_{\mathcal{K},\alpha/2}$, $\alpha \in (0,1)$ for appropriately chosen grids. The tables in particular show that in all considered cases the order of the remainder term $r_{\mathcal{K},a_\bullet}^2(\delta^2 n)$ is negligible compared with $\rho_{\mathcal{K},a_\bullet}^2(\delta n)$. In a mildly ill-posed model (parameter p) with ordinary smoothness (parameter s) we have seen in Illustration 4.2.5 that the optimal dimension $k_{a_\bullet}$ is of order $n^{\frac{2}{4p+4s+1}}$, which is smaller than n^2 for all combinations of s and p, by the reasoning of Remark 4.3.5 it is even smaller than $n^2/2$. Hence, we choose the geometric grid

$$\mathcal{K}_2 = \left\{2^j, j \in \left\{0, \dots, \lfloor \log_2(\tfrac{n^2}{2}) \rfloor\right\}\right\} \qquad \text{with} \qquad |\mathcal{K}_2| = \lfloor \log_2(\tfrac{n^2}{2}) \rfloor$$

and obtain the adaptive factor $\delta^{-1} = (1 + \log|\mathcal{K}_2|)^{1/2} \sim \sqrt{\log\log n}$. It is easily seen that the remainder term $r_{\mathcal{K}_2,a_\bullet}^2(\delta^2 n)$ is asymptotically negligible compared with $\rho_{\mathcal{K}_2,a_\bullet}^2(\delta n)$, since for some positive constants $x, y > 0$ (depending on s and p) we have

$$\frac{r_{\mathcal{K}_2,a_\bullet}^2(\delta^2 n)}{\rho_{\mathcal{K}_2,a_\bullet}^2(\delta n)} \sim \frac{(\log\log n)^x}{n^y} \longrightarrow 0 \qquad (n \to \infty).$$

Moreover, $\frac{\rho_{\mathcal{K}_2,a_\bullet}^2(\delta n)}{\delta^3}$ tends to zero for $n \to \infty$, since δ is only of log-order. Therefore, the upper bound derived in Proposition 4.3.2 asymptotically reduces to $\rho_{\mathcal{K}_2,a_\bullet}^2(\delta n)$, which is of the same order as $\rho_{a_\bullet}^2(\delta n)$ with an adaptive factor of order $\sqrt{\log\log n}$.

Order of $r_{\mathcal{K}_2,a_\bullet}^2(\delta^2 n)$ and $\rho_{\mathcal{K}_2,a_\bullet}^2(\delta n)$ with $\delta = (1 + \log|\mathcal{K}_2|)^{-1/2}$ and $\mathcal{K}_2 = \left\{2^j, j \in \left\{0, \dots, \lfloor \log_2(\tfrac{n^2}{2}) \rfloor\right\}\right\}$

a_j (smoothness)	$\lvert\varphi_j\rvert$ (ill-posedness)	$r_{\mathcal{K}_2,a_\bullet}^2(\delta^2 n)$	$\rho_{\mathcal{K}_2,a_\bullet}^2(\delta n)$
j^{-s}	$\lvert j\rvert^{-p}$	$\left(\frac{n}{\log\log n}\right)^{-\frac{4s}{4s+4p}}$	$\left(\frac{n}{(\log\log n)^{1/2}}\right)^{-\frac{4s}{4s+4p+1}}$

In a severely ill-posed model with ordinary smoothness, we have seen in Illustration 4.2.5 that the order of the optimal dimension parameter does not depend on the smoothness parameter $a_\bullet$. Hence, the test $\Delta_{\alpha/2,k_{a_\bullet}}$, which is, in fact, independent of $a_\bullet$, is automatically adaptive with respect to ordinary smoothness and no aggregation procedure is needed.

In a mildly ill-posed model with super smoothness (parameter s) Illustration 4.2.5 shows that the optimal dimension is of order $(\log n)^{\frac{1}{s}}$. Hence, we choose the smaller geometric grid

$$\mathcal{K}_{s_\star} = \left\{2^j, j \in \left\{0, \dots, \lfloor \tfrac{\log_2 \log n}{s_\star} \rfloor\right\}\right\} \qquad \text{with} \qquad |\mathcal{K}_{s_\star}| = \lfloor \tfrac{\log_2 \log n}{s_\star} \rfloor$$

for adaptation to $s \geq s_\star$ and obtain an even smaller adaptive factor $\delta^{-1} = (1+\log|\mathcal{K}_{s_\star}|)^{1/2} \sim \sqrt{\log\log\log n}$. It is easily seen that the remainder term $r^2_{\mathcal{K}_{s_\star},a_\bullet}(\delta^2 n)$ is asymptotically negligible compared with $\rho^2_{\mathcal{K}_{s_\star},a_\bullet}(\delta n)$, since

$$\frac{r^2_{\mathcal{K}_{s_\star},a_\bullet}(\delta^2 n)}{\rho^2_{\mathcal{K}_{s_\star},a_\bullet}(\delta n)} \sim \frac{\sqrt{\log\log\log n}}{(\log n)^{\frac{1}{2s}}} \longrightarrow 0 \qquad (n \to \infty).$$

Moreover, $\frac{\rho^2_{\mathcal{K}_{s_\star},a_\bullet}(\delta n)}{\delta^3}$ tends to zero for $n \to \infty$, since δ is only of log-order. Therefore, the upper bound derived in Proposition 4.3.2 asymptotically reduces to $\rho^2_{\mathcal{K}_{s_\star},a_\bullet}(\delta n)$, which is of the same order as $\rho^2_{a_\bullet}(\delta n)$ with an adaptive factor of order $\sqrt{\log\log\log n}$.

Order of $r^2_{\mathcal{K}_{s_\star},a_\bullet}(\delta^2 n)$ and $\rho^2_{\mathcal{K}_{s_\star},a_\bullet}(\delta n)$ with $\delta = (1+\log|\mathcal{K}_{s_\star}|)^{-1/2}$ and $\mathcal{K}_{s_\star} = \left\{2^j, j \in \left\{0,\dots,\lfloor \frac{\log_2 \log n}{s_\star} \rfloor\right\}\right\}$			
a_j (smoothness)	$\lvert\varphi_j\rvert$ (ill-posedness)	$r^2_{\mathcal{K}_{s_\star},a_\bullet}(\delta^2 n)$	$\rho^2_{\mathcal{K}_{s_\star},a_\bullet}(\delta n)$
e^{-j^s}	$\lvert j\rvert^{-p}$	$\frac{\log\log\log n}{n}(\log n)^{\frac{2p}{s}}$	$\frac{(\log\log\log n)^{1/2}}{n}(\log n)^{\frac{2p+1/2}{s}}$

Calculations for the risk bounds in Illustration 4.3.6.
Firstly, we determine the order of the terms $\rho^2_{\mathcal{K},a_\bullet}(\delta n) = \min_{k\in\mathcal{K}}\left\{a_k^2 \vee \frac{\nu_k^2}{\delta n}\right\}$.

1. **(ordinary smooth - mildly ill-posed)**
We first show that minimisation over $\mathcal{K}_2$ approximates the minimisation over $\mathbb{N}$ well enough, i.e.

$$\rho^2_{\mathcal{K}_2,a_\bullet}(n) = \min_{k\in\mathcal{K}_2} \rho^2_{k,a_\bullet}(n) \sim \min_{k\in\mathbb{N}} \rho^2_{a_\bullet}(n) = \rho^2_{a_\bullet}(n).$$

We aim to find $j_\star$ such that $2^{j_\star} \in \mathcal{K}_2$ approximates $k_{a_\bullet}$ well. Since $k_{a_\bullet} \sim n^{\frac{2}{4p+4s+1}}$ we define

$$j_\star := \lceil \frac{2}{4p+4s+1} \log_2 n \rceil \lesssim \lceil \log(n^2/2) \rceil.$$

Then,

$$\begin{aligned}\rho^2_{\mathcal{K}_2,a_\bullet}(n) \le \rho^2_{2^{j_\star},a_\bullet}(n) = a^2_{2^{j_\star}} \vee \frac{\nu^2_{2^{j_\star}}}{n} &\lesssim 2^{-2sj_\star} \vee \frac{2^{j_\star(2p+1/2)}}{n} \\ &\lesssim 2^{-(\frac{4s}{4p+4s+1}\log_2 n)} \vee \frac{2^{(\frac{2}{4p+4s+1}\log_2 n+1)(2p+1/2)}}{n} \\ &\lesssim n^{-\frac{4s}{4p+4s+1}} \vee 2^{2p+1/2} n^{-\frac{4s}{4p+4s+1}} \\ &\lesssim n^{-\frac{4s}{4p+4s+1}} \sim \rho^2_{a_\bullet}(n)\end{aligned}$$

Since, trivially $\rho^2_{a_\bullet}(n) \le \rho^2_{\mathcal{K}_2,a_\bullet}(n)$, we obtain $\rho^2_{a_\bullet}(n) \sim \rho^2_{\mathcal{K}_2,a_\bullet}(n)$. Replacing n by δn yields the result.

2. **(super smooth - mildly ill-posed)**
We first show that minimisation over $\mathcal{K}_{s_\star}$ approximates the minimisation over $\mathbb{N}$ well enough, i.e.

$$\rho^2_{\mathcal{K}_{s_\star},a_\bullet}(n) = \min_{k\in\mathcal{K}_{s_\star}} \rho^2_{k,a_\bullet}(n) \sim \min_{k\in\mathbb{N}} \rho^2_{a_\bullet}(n) = \rho^2_{a_\bullet}(n).$$

We aim to find $j_\star$ such that $2^{j_\star} \in \mathcal{K}_{s_\star}$ approximates $k_{a_\bullet}$ well. Since $k_{a_\bullet} \sim (\log n)^{\frac{1}{s}}$ we define

$$j_\star := \lceil \frac{1}{s}\log_2 \log n \rceil \lesssim \lceil \frac{1}{s_\star} \log\log(n) \rceil.$$

Then,

$$\begin{aligned}\rho^2_{\mathcal{K}_{s_\star},a_\bullet}(n) \le \rho^2_{2^{j_\star},a_\bullet}(n) = a^2_{2^{j_\star}} \vee \frac{\nu^2_{2^{j_\star}}}{n} &\lesssim e^{-2\cdot 2^{sj_\star}} \vee \frac{2^{j_\star(2p+1/2)}}{n} \\ &\lesssim e^{-2\cdot 2^{s\frac{1}{s}\log_2\log n}} \vee \frac{2^{(\frac{1}{s}\log_2\log n+1)(2p+1/2)}}{n} \\ &\lesssim n^{-2} \vee 2^{2p+1/2}\frac{(\log n)^{\frac{2p+1/2}{s}}}{n} \\ &\lesssim \frac{(\log n)^{\frac{2p+1/2}{s}}}{n} \sim \rho^2_{a_\bullet}(n)\end{aligned}$$

Since, trivially $\rho^2_{a_\bullet}(n) \le \rho^2_{\mathcal{K}_{s_\star},a_\bullet}(n)$, we obtain $\rho^2_{a_\bullet}(n) \sim \rho^2_{\mathcal{K}_{s_\star},a_\bullet}(n)$. Replacing n by $\delta^2 n$ yields the result.

Next, we determine the order of the remainder term $r^2_{\mathcal{K},a_\bullet}(\delta^2 n) = \min_{k\in\mathcal{K}}\left\{a^2_k \vee \frac{m^2_k}{\delta^2 n}\right\}$, by first calculating $r^2_{\mathbb{N},a_\bullet}(n) = \min_{k\in\mathbb{N}}\left\{a^2_k \vee \frac{m^2_k}{n}\right\}$ and then showing that minimisation over $\mathcal{K}$ approximates the minimisation over $\mathbb{N}$ well enough.

1. **(ordinary smooth - mildly ill-posed)** The variance term $\frac{m^2_k}{n}$ is of order $\frac{k^{2p}}{n}$ and the bias term a^2_k is of order k^{-2s}. Hence, the minimizing $k_\star$ satisfies $k_\star^{-2s} \sim \frac{k_\star^{2p}}{n}$ and thus $k_\star \sim n^{\frac{1}{2s+2p}}$, which yields $r^2_{\mathbb{N},a_\bullet}(n) \sim n^{-\frac{s}{s+p}}$.
Next, we show that minimisation over $\mathcal{K}_2$ approximates the minimization over $\mathbb{N}$ well enough, i.e.

$$r^2_{\mathcal{K}_2,a_\bullet}(n) = \min_{k\in\mathcal{K}_2} r^2_{k,a_\bullet}(n) \sim \min_{k\in\mathbb{N}} r^2_{k,a_\bullet}(n) = r^2_{a_\bullet}(n)$$

We aim to find $j_\star$ such that $2^{j_\star} \in \mathcal{K}_2$ approximates $k_\star$ well. Since $k_\star \sim n^{\frac{1}{2s+2p}}$ we define

$$j_\star := \lceil \frac{1}{2s+2p} \log_2 n \rceil \lesssim \lceil \log(n^2/2) \rceil.$$

Then,

$$\begin{aligned} r^2_{\mathcal{K}_2, a_\bullet}(n) \le r^2_{2^{j_\star}, a_\bullet}(n) = a^2_{2^{j_\star}} \vee \frac{m^2_{2^{j_\star}}}{n} &\lesssim 2^{-2sj_\star} \vee \frac{2^{j_\star(2p)}}{n} \\ &\lesssim 2^{-(\frac{2s}{2p+2s}\log_2 n)} \vee \frac{2^{(\frac{1}{2p+2s}\log_2 n+1)(2p)}}{n} \\ &\lesssim n^{-\frac{s}{p+s}} \vee 2^{2p} n^{-\frac{s}{p+s}} \\ &\lesssim n^{-\frac{s}{p+s}} \sim r^2_{\mathbb{N}, a_\bullet}(n) \end{aligned}$$

Since, trivially $r^2_{\mathbb{N},a_\bullet}(n) \le r^2_{\mathcal{K}_2,a_\bullet}(n)$, we obtain $r^2_{\mathbb{N},a_\bullet}(n) \sim r^2_{\mathcal{K}_2,a_\bullet}(n)$. Replacing n by δn yields the result.

2. **(super smooth - mildly ill-posed)** The variance term $\frac{m_k^2}{n}$ is of order $\frac{k^{2p}}{n}$ and the bias term a_k^2 is of order e^{-2k^s}. Hence, the minimizing $k_\star$ satisfies $e^{-2k_\star^s} \sim \frac{k_\star^{2p}}{n}$ and thus $k_\star \sim (\log(n/b_n))^{\frac{1}{s}}$ with $b_n \sim (\log n)^{\frac{2p}{s}}$, which yields $r^2_{\mathbb{N},a_\bullet}(n) \sim \frac{(\log n)^{\frac{2s}{p}}}{n}$.
Next, we show that minimisation over $\mathcal{K}_{s_\star}$ approximates the minimization over $\mathbb{N}$ well enough, i.e.

$$r^2_{\mathcal{K}_{s_\star}, a_\bullet}(n) = \min_{k \in \mathcal{K}_{s_\star}} r^2_{k,a_\bullet}(n) \sim \min_{k \in \mathbb{N}} r^2_{k,a_\bullet}(n) = r^2_{a_\bullet}(n)$$

We aim to find $j_\star$ such that $2^{j_\star} \in \mathcal{K}_{s_\star}$ approximates $k_\star$ well. Since $k_\star \sim (\log n)^{\frac{1}{s}}$ we define

$$j_\star := \lceil \frac{1}{s} \log_2 \log n \rceil \lesssim \lceil \frac{1}{s_\star} \log \log n \rceil.$$

Then,

$$\begin{aligned} r^2_{\mathcal{K}_{s_\star}, a_\bullet}(n) \le r^2_{2^{j_\star}, a_\bullet}(n) = a^2_{2^{j_\star}} \vee \frac{m^2_{2^{j_\star}}}{n} &\lesssim e^{-2\cdot 2^{sj_\star}} \vee \frac{2^{j_\star(2p)}}{n} \\ &\lesssim e^{-2\cdot 2^{s\frac{1}{s}\log_2 \log n}} \vee \frac{2^{(\frac{1}{s}\log_2 \log n+1)(2p)}}{n} \\ &\lesssim n^{-2} \vee 2^{2p+1/2} \frac{(\log n)^{\frac{2p}{s}}}{n} \\ &\lesssim \frac{(\log n)^{\frac{2p}{s}}}{n} \sim r^2_{\mathbb{N},a_\bullet}(n) \end{aligned}$$

Since, trivially $r^2_{\mathbb{N},a_\bullet}(n) \le r^2_{\mathcal{K}_{s_\star},a_\bullet}(n)$, we obtain $r^2_{\mathbb{N},a_\bullet}(n) \sim r^2_{\mathcal{K}_{s_\star},a_\bullet}(n)$. Replacing n by $\delta^2 n$ yields the result.

□

Remark 4.3.7 (Adaptation to the radius R of the alternative). *In this chapter the parameter* R *is unknown but assumed to be fixed and we consider adaptation to a collection of alternatives* $\left\{\mathcal{E}^{\mathrm{R}}_{a_\bullet} : a_\bullet \in \mathcal{A}\right\}$ *only. From* Corollary 4.2.4 *(and the definition of* $\overline{A}_\alpha$ *therein) it follows immediately that adaptation to* $\left\{\mathcal{E}^{\mathrm{R}}_{a_\bullet} : \mathrm{R} \in (0, \mathrm{R}^\star]\right\}$ *is achieved without a loss. Indeed, replacing* R *by* $\mathrm{R}^\star$ *in the definition of* $\overline{A}_\alpha$ *we promptly obtain a result similar to* (4.2.19) *in* Corollary 4.2.4 *with an additional supremum taken over* $\mathrm{R} \in (0, \mathrm{R}^\star]$*. However, adaptation to* $\left\{\mathcal{E}^{\mathrm{R}}_{a_\bullet} : \mathrm{R} \in (0, \infty)\right\}$ *is not possible without a loss, for an explanation of this phenomenon we refer to Section 6.3. in Baraud [2002] for a similar observation in a Gaussian sequence space model.* □

4.4 Upper bound via a direct testing procedure

Definition of the test statistic. Instead of estimating the $\mathscr{L}^2$-distance $\mathrm{q}^2(f - f^\circ)$, in this section we consider a test that is based on an estimation of the $\mathscr{L}^2$-distance between the images $g = f \circledast \varphi$ and $g^\circ = f^\circ \circledast \varphi$, i.e. of

$$\mathrm{q}^2(g - g^\circ) = \sum_{|j| \in \mathbb{N}} \left| g_j - g_j^\circ \right|^2 = \sum_{|j| \in \mathbb{N}} |f - f^\circ|^2 \, |\varphi_j|^2 .$$

For $k \in \mathbb{N}$ we construct an unbiased estimator $\widetilde{\mathrm{q}}_k^2$ of the truncated version

$$\mathrm{q}_k^2(g - g^\circ) := \sum_{|j| \in [\![k]\!]} \left| g_j - g_j^\circ \right|^2 = \sum_{|j| \in [\![k]\!]} |g_j|^2 - 2 \sum_{|j| \in [\![k]\!]} g_j^\circ \overline{g_j} + \sum_{|j| \in [\![k]\!]} \left| g_j^\circ \right|^2 . \tag{4.4.1}$$

The first two summands of (4.4.1) are unknown and, thus, need to be estimated, the third is known. For the second term, which is a linear term, we plug in canonical estimators of the Fourier coefficients $\overline{g_j}$ and obtain

$$\tilde{S}_k := \frac{1}{n} \sum_{|j| \in [\![k]\!]} \sum_{l \in [\![n]\!]} g_j^\circ e_j(Y_l)$$

as an unbiased estimator of $\sum_{|j| \in [\![k]\!]} g_j^\circ \overline{g_j}$. For the first term, which is quadratic, we use the U-statistic

$$\tilde{T}_k = \frac{1}{n(n-1)} \sum_{|j| \in [\![k]\!]} \sum_{\substack{l,m \in [\![k]\!] \\ l \neq m}} e_j(-Y_l) e_j(Y_m)$$

as an unbiased estimator of $\sum_{|j| \in [\![k]\!]} |g_j|^2$. In total, we consider the test statistic

$$\widetilde{\mathrm{q}}_k^2 := \tilde{T}_k - 2\tilde{S}_k + \mathrm{q}_k^2(g^\circ).$$

Decomposition of the test statistic. Similarly to the decomposition (4.2.5) of the indirect test statistic, we split $\widetilde{\mathrm{q}}_k^2$ into three parts;

$$\widetilde{\mathrm{q}}_k^2 = \mathrm{U}_n^{\mathrm{d}} + 2\mathrm{V}_n^{\mathrm{d}} + \mathrm{q}_k^2(g - g^\circ) \tag{4.4.2}$$

with the canonical U-statistic

$$\mathrm{U}_n^{\mathrm{d}} := \frac{1}{n(n-1)} \sum_{|j| \in [\![k]\!]} \sum_{\substack{l,m \in [\![n]\!] \\ l \neq m}} (e_j(-Y_l) - g_j)(e_j(Y_m) - \overline{g}_j), \tag{4.4.3}$$

the centred linear term

$$\mathrm{V}_n^{\mathrm{d}} := \frac{1}{n} \sum_{|j| \in [\![k]\!]} \sum_{l \in [\![n]\!]} (g_j - g_j^\circ)(e_j(Y_l) - \overline{g_j}) \tag{4.4.4}$$

and the separation term $\mathrm{q}_k^2(g - g^\circ)$.

Definition of the threshold. In the next proposition we provide bounds for the quantiles of the test statistic $\widetilde{\mathrm{q}}_k^2$. Recall that $L_x := (\log(e/x))^{1/2} = (1 - \log(x))^{1/2} \in (1, \infty)$ for $x \in (0,1)$. For $c_1 := 799 \, \|g_\bullet^\circ\|_{\ell^2} + 1372$, $c_2 := 52 \, \|g_\bullet^\circ\|_{\ell^1}$ and $\alpha \in (0,1)$, we define the threshold

$$\tau_k^{\mathrm{d}}(\alpha) := c_1 \left(1 \vee L_\alpha^2 \sqrt{\frac{\sqrt{2k}}{n}} \vee L_\alpha^3 \frac{\sqrt{2k}}{n} \right) L_\alpha \frac{\sqrt{2k}}{n} + c_2 L_\alpha^2 \frac{1}{n}. \tag{4.4.5}$$

Proposition 4.4.1 (Bounds for the quantiles of $\widetilde{\mathrm{q}}_k^2$). For densities $f^\circ, f, \varphi \in \mathcal{D}$ and $n \in \mathbb{N}, n \geq 2$ consider $\{Y_j\}_{j=1}^n \overset{\text{iid}}{\sim} g = f \circledast \varphi$ with joint distribution $\mathbb{P}_f$ and let $g^\circ = f^\circ \circledast \varphi$. Let $\alpha, \beta \in (0,1)$ and for $k \in \mathbb{N}$ consider the estimator $\widetilde{\mathrm{q}}_k^2$ and the threshold $\tau_k^{\mathrm{d}}(\alpha)$ as defined in (4.4.2) and (4.4.5), respectively.

(i) If $\mathscr{L}^2(g^\circ) = \{|\xi|, \xi \in \mathscr{L}^2\}$, then

$$\mathbb{P}_{f^\circ}(\widetilde{\mathrm{q}}_k^2 \geq \tau_k^{\mathrm{d}}(\alpha)) \leq \alpha.$$

(ii) If $c_3 := 837 \left\| \varphi_\bullet \right\|_{\ell^2} + 1373$ and the separation condition

$$\mathrm{q}_k^2(g - g^\circ) \geq 2 \left(\tau_k^{\mathrm{d}}(\alpha) + c_3 L_{\beta/2}^4 \left(1 \vee \frac{\sqrt{2k}}{n} \right) \frac{\sqrt{2k}}{n} \right), \tag{4.4.6}$$

holds, then

$$\mathbb{P}_f(\widetilde{\mathrm{q}}_k^2 < \tau_k^{\mathrm{d}}(\alpha)) \leq \beta.$$

Proof of Proposition 4.4.1. The proof is similar to the proof of Proposition 4.2.1 using the decomposition (4.4.2) rather than (4.2.5). For the first part (i), we apply Proposition D.1.1 of the appendix together with Lemma D.3.1 (instead of Lemma D.2.1). For the second part (ii), we control the deviations of the U-statistic $\mathrm{U}_n^{\mathrm{d}}$ and the linear statistic $\mathrm{V}_n^{\mathrm{d}}$ by applying Lemma D.3.1 and Lemma D.3.2 (instead of Lemma D.2.1 and Lemma D.2.2.)

(i) If $f = f^\circ$ and, hence, $g = g^\circ$, the decomposition (4.4.2) simplifies to $\widetilde{\mathrm{q}}_k^2 = \mathrm{U}_n^{\mathrm{d}}$, where $\mathrm{U}_n^{\mathrm{d}}$ is a canonical U-statistic. Applying Proposition D.1.1 of the appendix, a concentration inequality for canonical U-statistics of order 2, with $x = L_\alpha^2 \geq 1$ and quantities $A - D$ satisfying (D.1.2), we obtain

$$\mathbb{P}_{f^\circ}\left(\mathrm{U}_n^{\mathrm{d}} \geq 8 \frac{C}{n} L_\alpha + 13 \frac{D}{n} L_\alpha^2 + 261 \frac{B}{n^{3/2}} L_\alpha^3 + 343 \frac{A}{n^2} L_\alpha^4 \right) \leq \exp(1 - x). \tag{4.4.7}$$

Consider the quantities $A - C$ defined in (D.3.1) and D in (D.3.2), which under the additional assumption $\mathscr{L}^2(g^\circ) = \{|\xi| : \xi \in \mathscr{L}^2\}$ satisfy (D.1.2) due to Lemma D.3.1. We have

$$\begin{aligned}
&8 \frac{C}{n} L_\alpha + 13 \frac{D}{n} L_\alpha^2 + 261 \frac{B}{n^{3/2}} L_\alpha^3 + 343 \frac{A}{n^2} L_\alpha^4 \\
\leq & 8 \cdot 2 \cdot \left\| g_\bullet \right\|_{\ell^2} L_\alpha \frac{\sqrt{2k}}{n} + 13 \cdot 4 \cdot \left\| g_\bullet \right\|_{\ell^1} L_\alpha^2 \frac{1}{n} \\
&+ 261 \cdot 3 \cdot \left\| g_\bullet \right\|_{\ell^2} L_\alpha^3 \frac{(2k)^{3/2}}{n^{3/2}} + 343 \cdot 4 \cdot L_\alpha^4 \frac{2k}{n^2} \\
= & 52 \left\| g_\bullet \right\|_{\ell^1} L_\alpha^2 \frac{1}{n} + L_\alpha \frac{\sqrt{2k}}{n} \left(16 \left\| g_\bullet \right\|_{\ell^2} + 783 \left\| g_\bullet \right\|_{\ell^2} L_\alpha^2 \frac{(2k)^{1/4}}{n^{1/2}} + 1372 L_\alpha^3 \frac{\sqrt{2k}}{n} \right) \\
\leq & 52 \left\| g_\bullet \right\|_{\ell^1} L_\alpha^2 \frac{m_k^2}{n} + L_\alpha \frac{\sqrt{2k}}{n} (799 \left\| g_\bullet \right\|_{\ell^2} + 1372) \left(1 \vee L_\alpha^2 \frac{(2k)^{1/4}}{n^{1/2}} \vee L_\alpha^3 \frac{\nu_k^2}{n} \right) \\
= & c_2 L_\alpha^2 \frac{1}{n} + c_1 \left(1 \vee L_\alpha^2 \sqrt{\frac{\sqrt{2k}}{n}} \vee L_\alpha^3 \frac{\sqrt{2k}}{n} \right) L_\alpha \frac{\sqrt{2k}}{n} \\
= & \tau_k^{\mathrm{d}}(\alpha),
\end{aligned}$$

which together with (4.2.11) shows the assertion (i).

(ii) Keeping the decomposition (4.4.2) in mind, we control the deviations of the U-statistic $\mathrm{U}_n^{\mathrm{d}}$ and the linear statistic $\mathrm{V}_n^{\mathrm{d}}$ by applying Proposition D.1.1 and Lemma D.3.2 of the appendix, respectively. In fact, the quantities $A - D$ given in (D.2.1) of Lemma D.3.1 fulfil (recall that $L_{\beta/2} \geq 1$ for all $\beta > 0$)

$$\begin{aligned}
&8\frac{C}{n}L_{\beta/2} + 13\frac{D}{n}L_{\beta/2}^2 + 261\frac{B}{n^{3/2}}L_{\beta/2}^3 + 343\frac{A}{n^2}L_{\beta/2}^4 \\
\leq & 8 \cdot 2 \cdot \|g_\bullet\|_{\ell^2} L_{\beta/2}\frac{\sqrt{2k}}{n} + 13 \cdot 2 \cdot \|g_\bullet\|_{\ell^2} L_{\beta/2}^2\frac{\sqrt{2k}}{n} \\
&\qquad + 261 \cdot 3 \cdot \|g_\bullet\|_{\ell^2} L_{\beta/2}^3\frac{(2k)^{3/2}}{n^{3/2}} + 343 \cdot 4 \cdot L_{\beta/2}^4\frac{2k}{n^2} \\
\leq & L_{\beta/2}^4\left(42\,\|g_\bullet\|_{\ell^2} + 783\,\|g_\bullet\|_{\ell^2}\frac{(2k)^{1/4}}{n^{1/2}} + 1372\frac{\sqrt{2k}}{n}\right)\frac{\sqrt{2k}}{n} \\
\leq & L_{\beta/2}^4\,(825\,\|g_\bullet\|_{\ell^2} + 1372)\left(1 \vee \frac{(2k)^{1/4}}{n^{1/2}} \vee \frac{\sqrt{2k}}{n}\right)\frac{\sqrt{2k}}{n} \\
\leq & L_{\beta/2}^4(825\,\|g_\bullet\|_{\ell^2} + 1372)\left(1 \vee \frac{\sqrt{2k}}{n}\right)\frac{\sqrt{2k}}{n} =: \tau_1^{\mathrm{d}},
\end{aligned}$$

where we exploited that $1 \vee a \vee a^2 = 1 \vee a^2$ for any $a \geq 0$. Consequently, the event

$$\Omega_1^{\mathrm{d}} := \left\{\mathrm{U}_n^{\mathrm{d}} \leq -\tau_1^{\mathrm{d}}\right\}$$

satisfies $\mathbb{P}_f(\Omega_1^{\mathrm{d}}) \leq \beta/2$ due to Proposition D.1.1 (with the usual symmetry argument). Define further the event

$$\Omega_2^{\mathrm{d}} := \left\{2\mathrm{V}_n^{\mathrm{d}} \leq -\tau_2^{\mathrm{d}} - \frac{1}{2}\mathrm{q}_k^2(g - g^\circ)\right\}$$

with $\tau_2^{\mathrm{d}} := L_{\beta/2}^2\,(12\,\|g_\bullet^\circ\|_{\ell^2} + 1)\left(1 \vee \frac{\sqrt{2k}}{n}\right)\frac{\sqrt{2k}}{n}$. Then we have $\mathbb{P}_f(\Omega_2) \leq \frac{\beta}{2e} \leq \frac{\beta}{2}$ due to Lemma D.3.2 with $x = L_{\beta/2} \geq 1$, which is an application of a Bernstein-type inequality. We obtain

$$\begin{aligned}
\tau_1^{\mathrm{d}} + \tau_2^{\mathrm{d}} &= L_{\beta/2}^4(825\,\|g_\bullet\|_{\ell^2} + 1372)\left(1 \vee \frac{\sqrt{2k}}{n}\right)\frac{\sqrt{2k}}{n} \\
&\qquad + L_{\beta/2}^2(12\,\|g_\bullet\|_{\ell^2} + 1)\left(1 \vee \frac{\sqrt{2k}}{n}\right)\frac{\sqrt{2k}}{n} \\
&\leq L_{\beta/2}^4\left(1 \vee \frac{\sqrt{2k}}{n}\right)\frac{\sqrt{2k}}{n}\,(825\,\|g_\bullet\|_{\ell^2} + 1372 + 12\,\|g_\bullet\|_{\ell^2} + 1) \\
&\leq L_{\beta/2}^4\left(1 \vee \frac{\sqrt{2k}}{n}\right)\frac{\sqrt{2k}}{n}\,(837\,\|g_\bullet\|_{\ell^2} + 1373)
\end{aligned}$$

with $c_3 = 837\,\|\varphi_\bullet\|_{\ell^2} + 1373$ due to $1 \leq L_{\beta/2}$ and $\|g_\bullet\|_{\ell^2} \leq \|\varphi_\bullet\|_{\ell^2}$. Hence, the assumption (4.4.6) implies

$$\frac{1}{2}\mathrm{q}_k^2(g - g^\circ) \geq \tau_k^{\mathrm{d}}(\alpha) + \tau_1^{\mathrm{d}} + \tau_2^{\mathrm{d}}.$$

The decomposition (4.4.2) yields

$$\begin{aligned}
\mathbb{P}_f\left(\widetilde{\mathrm{q}}_k^2 < \tau_k^{\mathrm{d}}(\alpha)\right) &= \mathbb{P}_f\left(\left\{\widetilde{\mathrm{q}}_k^2 < \tau_k(\alpha)\right\} \cap \Omega_1^{\mathrm{d}}\right) + \mathbb{P}_f\left(\left\{\hat{\mathrm{q}}_k^2 < \tau_k(\alpha)\right\} \cap (\Omega_1^{\mathrm{d}})^c\right) \\
&\leq \mathbb{P}_f(\Omega_1^{\mathrm{d}}) + \mathbb{P}_f\left(2\mathrm{V}_n^{\mathrm{d}} + \mathrm{q}_k^2(g - g^\circ) < \tau_k^{\mathrm{d}}(\alpha) + \tau_1^{\mathrm{d}}\right) \\
&\leq \frac{\beta}{2} + \mathbb{P}_f(\Omega_2^{\mathrm{d}}) \leq \beta,
\end{aligned}$$

which shows (ii) and completes the proof.

□

Definition of the test. For $k \in \mathbb{N}$ and $\alpha \in (0,1)$ we define the test

$$\Delta^{\mathrm{d}}_{k,\alpha} := \mathbb{1}_{\{\widetilde{\mathrm{q}}^2_k \geq \tau^{\mathrm{d}}_k(\alpha)\}}, \tag{4.4.8}$$

where $\widetilde{\mathrm{q}}^2_k$ is the test statistic in (4.4.2) and $\tau^{\mathrm{d}}_k(\alpha)$ the threshold in (4.4.5). Proposition 4.4.1 (i) shows that $\Delta^{\mathrm{d}}_{k,\alpha}$ is a level-α-test for all $k \in \mathbb{N}$. Moreover, Proposition 4.4.1 (ii) characterises elements for which $\Delta^{\mathrm{d}}_{k,\alpha}$ is $(1-\beta)$-powerful. Exploiting these results and considering the additional regularity constraint $\mathcal{E}^{\mathrm{R}}_{a_\bullet}$ (defined in (4.2.13)), we derive an upper bound for the radius of testing of $\Delta^{\mathrm{d}}_{k,\alpha}$ in terms of $m^2_k = \max_{|j| \in [\![k]\!]} |\varphi_j|^{-2}$ as in (4.2.8) and the regularity parameter $a_\bullet$, that is we define

$$(\rho^{\mathrm{d}}_{k,a_\bullet})^2 := (\rho^{\mathrm{d}}_{k,a_\bullet}(n))^2 := \left\{ a^2_k \vee \frac{\sqrt{2k}}{n} m^2_k \right\}. \tag{4.4.9}$$

Proposition 4.4.2 (Upper bound for the radius of testing of $\Delta^{\mathrm{d}}_{k,\alpha/2}$). Let $g^\circ = f^\circ \circledast \varphi$ with $f^\circ, \varphi \in \mathcal{D}$ satisfy $\mathscr{L}^2(g^\circ) = \{|\xi|, \xi \in \mathscr{L}^2\}$. For $\alpha \in (0,1)$ define

$$\overline{A}^2_\alpha := \mathrm{R}^2 + 2(837 \|\varphi_\bullet\|_{\ell^2} + 851 \|g^\circ_\bullet\|_{\ell^1} + 2745) L^4_{\alpha/4}. \tag{4.4.10}$$

For all $A \geq \overline{A}_\alpha$ and for all $n, k \in \mathbb{N}$ with $n \geq 2$ and $\sqrt{2k} \leq n$, we have

$$\mathcal{R}\left(\Delta^{\mathrm{d}}_{k,\alpha/2} \mid \mathcal{E}^{\mathrm{R}}_{a_\bullet}, A\rho^{\mathrm{d}}_{k,a_\bullet}\right) \leq \alpha.$$

Proof of Proposition 4.4.2. Using Proposition 4.4.1 we show that both the type I and the maximal type II error probability are bounded by $\alpha/2$, and thus the result follows from

$$\begin{aligned}\mathcal{R}\left(\Delta^{\mathrm{d}}_{k,\alpha/2} \mid \mathcal{E}^{\mathrm{R}}_{a_\bullet}, A\rho^{\mathrm{d}}_{k,a_\bullet}\right) &= \mathbb{P}_{f^\circ}\left(\Delta^{\mathrm{d}}_{k,\alpha/2} = 1\right) + \sup_{f - f^\circ \in \mathscr{L}^2_{A\rho^{\mathrm{d}}_{k,a_\bullet}} \cap \mathcal{E}^{\mathrm{R}}_{a_\bullet}} \mathbb{P}_f\left(\Delta^{\mathrm{d}}_{k,\alpha/2} = 0\right) \\ &\leq \alpha/2 + \alpha/2 = \alpha.\end{aligned}$$

Since the assumption of Proposition 4.4.1(i) is fulfilled, the test $\Delta^{\mathrm{d}}_{k,\alpha/2}$ is a level-$\alpha/2$-test. Hence, for each density $f \in \mathscr{L}^2$ belonging to the alternative, i.e. with $\|f - f^\circ\|^2_{\mathscr{L}^2} \geq \overline{A}^2_\alpha (\rho^{\mathrm{d}}_{k,a_\bullet})^2$ and $f - f^\circ \in \mathcal{E}^{\mathrm{R}}_{a_\bullet}$ it remains to verify condition (4.4.6) in order to apply Proposition 4.4.1 (ii) (with $\beta = \alpha/2$), i.e. we need to check that

$$\mathrm{q}^2_k(g - g^\circ) \geq 2\left(\tau^{\mathrm{d}}_k(\alpha/2) + c_3 L^4_{\alpha/4}\left(1 \vee \frac{\sqrt{2k}}{n}\right)\frac{\sqrt{2k}}{n}\right). \tag{4.4.11}$$

Indeed, in this situations we have

$$\sum_{|j|>k} \left|f_j - f^\circ_j\right|^2 \leq \sum_{|j|>k} \frac{a^2_k}{a^2_j}\left|f_j - f^\circ_j\right|^2 \leq a^2_k \mathrm{R}^2,$$

since $a_\bullet$ is non-increasing, hence $\frac{a_k}{a_j} \geq 1$ for all $j \geq k$. Therefore,

$$\begin{aligned}\mathrm{q}^2_k(f - f^\circ) &= \|f - f^\circ\|^2_{\mathscr{L}^2} - \sum_{|j|>k} |f - f^\circ|^2 \geq \overline{A}^2_\alpha (\rho^{\mathrm{d}}_{k,a_\bullet})^2 - a^2_k \mathrm{R}^2 \\ &\geq 2(837 \|\varphi_\bullet\|_{\ell^2} + 851 \|g^\circ_\bullet\|_{\ell^1} + 2745) L^4_{\alpha/4} \frac{\sqrt{2k}}{n} m^2_k,\end{aligned} \tag{4.4.12}$$

where the last inequality is due to the definition of $\overline{A}_\alpha$. Using $\|g^\circ_\bullet\|_{\ell^2} \leq \|g^\circ_\bullet\|_{\ell^1}$ we further obtain

$$\begin{aligned} \mathrm{q}_k^2(f - f^\circ) &\geq 2(837\,\|\varphi_\bullet\|_{\ell^2} + 799\,\|g^\circ_\bullet\|_{\ell^2} + 52\,\|g^\circ_\bullet\|_{\ell^1} + 2745)L^4_{\alpha/4}\frac{\sqrt{2k}}{n}m_k^2 \\ &\geq 2(c_1 + c_2 + c_3)L^4_{\alpha/4}\frac{\sqrt{2k}}{n}m_k^2. \end{aligned} \tag{4.4.13}$$

The condition (4.4.11) then follows from (4.4.13) by exploiting $1 \leq L_{\alpha/2} \leq L_{\alpha/4}$,$\|g^\circ_\bullet\|_{\ell^2} \leq \|g^\circ_\bullet\|_{\ell^1}$, $2k \leq n^2$ and $m_k^2\mathrm{q}_k^2(g - g^\circ) \geq \mathrm{q}_k^2(f - f^\circ)$, which holds since

$$\begin{aligned} \mathrm{q}_k^2(g - g^\circ)m_k^2 &= \sum_{|j| \in [\![k]\!]} \left|g_j - g_j^\circ\right|^2 \max_{|i| \in [\![k]\!]} |\varphi_i|^{-2} \\ &\geq \sum_{|j| \in [\![k]\!]} \frac{\left|g_j - g_j^\circ\right|^2}{|\varphi_j|^2} = \sum_{|j| \in [\![k]\!]} \left|f_j - f_j^\circ\right|^2 = \mathrm{q}_k^2(f - f^\circ). \end{aligned}$$

Indeed,

$$\begin{aligned} \tau_k^{\mathrm{d}}(\alpha/2) &= c_1\left(1 \vee L^2_{\alpha/2}\sqrt{\frac{\sqrt{2k}}{n}} \vee L^3_{\alpha/2}\frac{\sqrt{2k}}{n}\right)L_{\alpha/2}\frac{\sqrt{2k}}{n} + c_2L^2_{\alpha/2}\frac{1}{n} \\ &\leq (c_1 + c_2)L^4_{\alpha/4}\frac{\sqrt{2k}}{n}. \end{aligned}$$

Hence, due to $\sqrt{2k} \leq n$ and (4.4.13)

$$\begin{aligned} &2\left(\tau_k^{\mathrm{d}}(\alpha/2) + c_3L^4_{\alpha/4}\left(1 \vee \frac{\sqrt{2k}}{n}\right)\frac{\sqrt{2k}}{n}\right) \\ &\leq 2(c_1 + c_2 + c_3)L^4_{\alpha/4}\frac{\sqrt{2k}}{n} \\ &\leq \mathrm{q}_k^2(f - f^\circ)m_k^{-2} \leq \mathrm{q}_k^2(g - g^\circ), \end{aligned}$$

which completes the proof. □

The upper bound $(\rho^{\mathrm{d}}_{k,a_\bullet})^2$ for the radius of testing of $\Delta^{\mathrm{d}}_{k,\alpha/2}$ depends on the dimension parameter k. Let us introduce a dimension that realizes an optimal bias^2-variance trade-off and the corresponding radius

$$k^{\mathrm{d}}_{a_\bullet} := k^{\mathrm{d}}_{a_\bullet}(n) := \arg\min_{k \in \mathbb{N}} \rho^{\mathrm{d}}_{k,a_\bullet} := \min\left\{k \in \mathbb{N} : \rho^{\mathrm{d}}_{k,a_\bullet} \leq \rho^{\mathrm{d}}_{l,a_\bullet} \text{ for all } l \in \mathbb{N}\right\}$$

and

$$(\rho^{\mathrm{d}}_{a_\bullet})^2 := (\rho^{\mathrm{d}}_{a_\bullet}(n))^2 := \min_{k \in \mathbb{N}}(\rho^{\mathrm{d}}_{k,a_\bullet})^2 = \min_{k \in \mathbb{N}}\left\{a_k^2 \vee \frac{\sqrt{2k}}{n}m_k^2\right\}.$$

Corollary 4.4.3 (Upper bound for the radius of testing). Let $g^\circ = f^\circ \circledast \varphi$ with $f^\circ, \varphi \in \mathscr{L}^2$ satisfy $\mathscr{L}^2(g^\circ) = \{|\xi|, \xi \in \mathscr{L}^2\}$. For $\alpha \in (0, 1)$ define $\overline{A}_\alpha$ as in (4.4.10). Then for all $A \geq \overline{A}_\alpha$ and $n \geq \sqrt{2}\,|\varphi_1|^{-2}$, we have

$$\mathcal{R}\left(\Delta^{\mathrm{d}}_{k^{\mathrm{d}}_{a_\bullet},\alpha/2} \mid \mathcal{E}^{\mathrm{R}}_{a_\bullet}, A\rho^{\mathrm{d}}_{a_\bullet}\right) \leq \alpha.$$

Proof of Corollary 4.4.3. The result follows immediately from Proposition 4.4.2. Indeed, $n \geq \sqrt{2}\,|\varphi_1|^{-1}$ implies $1 \geq (\rho^{\mathrm{d}}_{1,a_\bullet})^2 \geq (\rho^{\mathrm{d}}_{a_\bullet})^2 \geq \frac{\sqrt{2k^{\mathrm{d}}_{a_\bullet}}}{n}$, which justifies the application of Proposition 4.4.2 and completes the proof. □

Remark 4.4.4 (Optimality of the direct testing procedure). *Let us compare the upper bound for the direct testing procedure* $(\rho^{\mathrm{d}}_{a_\bullet})^2 = \min_{k\in\mathbb{N}}\left\{a_k^2 \vee \frac{\sqrt{2k}}{n} m_k^2\right\}$ *with the minimax radius of testing* $\rho^2_{a_\bullet} = \min_{k\in\mathbb{N}}\left\{a_k^2 \vee \frac{\nu_k^2}{n}\right\}$. *Naturally,* $\rho^{\mathrm{d}}_{a_\bullet} \geq \rho_{a_\bullet}$. *Moreover, if there exists a constant* $c > 0$ *such that*

$$\nu_k^2 = \sqrt{\sum_{|j|\in[\![k]\!]} \frac{1}{|\varphi_j|^4}} \leq \sqrt{2k} \max_{|j|\in[\![k]\!]} |\varphi_j|^{-2} = \sqrt{2k}\, m_k^2 \leq c\nu_k^2, \tag{4.4.14}$$

then $\rho^{\mathrm{d}}_{a_\bullet}$ *and* $\rho_{a_\bullet}$ *are of the same order and, thus, the direct testing procedure is minimax optimal. Condition (4.4.14) is for instance satisfied for a mildly ill-posed model, i.e. if* $(|\varphi_j|)_{j\in\mathbb{N}}$ *decays polynomially. Note, however, that (4.4.14) is a sufficient but not a necessary condition. For a severely ill-posed model, i.e. if* $(|\varphi_j|)_{j\in\mathbb{N}}$ *decays exponentially, the condition (4.4.14) is not fulfilled. Nevertheless, the direct testing procedure still performs optimally (see Illustration 4.4.5 below).* □

Illustration 4.4.5. We illustrate the order of the upper bound for the radius of testing of the direct test $\Delta^{\mathrm{d}}_{k^{\mathrm{d}}_{a_\bullet},\alpha/2}$, $\alpha \in (0,1)$ under the regularity and ill-posedness assumptions introduced in Illustration 4.2.5. Comparing the resulting upper bounds $(\rho^{\mathrm{d}}_{a_\bullet})^2$ with the radii $\rho^2_{a_\bullet}$, we conclude that the direct test performs as well as the indirect test in all three cases.

Order of the optimal dimension $k^{\mathrm{d}}_{a_\bullet}$ and the upper bound $(\rho^{\mathrm{d}}_{a_\bullet})^2$			
a_j (smoothness)	$\lvert\varphi_j\rvert$ (ill-posedness)	$k^{\mathrm{d}}_{a_\bullet}$	$(\rho^{\mathrm{d}}_{a_\bullet})^2$
j^{-s}	$\lvert j\rvert^{-p}$	$n^{\frac{2}{4p+4s+1}}$	$n^{-\frac{4s}{4s+4p+1}}$
j^{-s}	$e^{-\lvert j\rvert^p}$	$(\log n)^{\frac{1}{p}}$	$(\log n)^{-\frac{2s}{p}}$
e^{-j^s}	$\lvert j\rvert^{-p}$	$(\log n)^{\frac{1}{s}}$	$n^{-1}(\log n)^{\frac{2p+1/2}{s}}$

Calculations for the radius bounds in Illustration 4.4.5.
Recall the definition $(\rho^{\mathrm{d}}_{a_\bullet})^2 = \min_{k\in\mathbb{N}}\left\{a_k^2 \vee \frac{\sqrt{2k}}{n} m_k^2\right\}$.

1. **(ordinary smooth - mildly ill-posed)** The variance term $m_k^2\frac{\sqrt{2k}}{n}$ is of order $\frac{1}{n}k^{2p+1/2}$ and the bias term a_k^2 is of order k^{-2s}. Hence, the optimal $k^{\mathrm{d}}_{a_\bullet}$ satisfies $(k_{a_\bullet})^{-2s} \sim \frac{1}{n}(k_{a_\bullet})^{2p+1/2}$ and thus $k_{a_\bullet} \sim n^{\frac{2}{4s+4p+1}}$, which yields an upper bound of order $(\rho^{\mathrm{d}}_{a_\bullet})^2 \sim (k_{a_\bullet})^{-2s} \sim n^{-\frac{4s}{4s+4p+1}}$.

2. **(ordinary smooth - severly ill-posed)** The variance term $m_k^2\frac{\sqrt{2k}}{n}$ is of order $\frac{1}{n}k^{1/2}e^{2k^p}$ and the bias term a_k^2 is of order k^{-2s}. Hence, the optimal $k^{\mathrm{d}}_{a_\bullet}$ satisfies $(k_{a_\bullet})^{-2s} \sim \frac{1}{n}(k_{a_\bullet})^{1/2}e^{2k^p}$ and thus $k^{\mathrm{d}}_{a_\bullet} \sim (\log(n/b_n))^{\frac{1}{p}}$ with $b_n \sim (\log n)^{\frac{4s+1}{2p}}$, which yields an upper bound of order $(\rho^{\mathrm{d}}_{a_\bullet})^2 \sim (k^{\mathrm{d}}_{a_\bullet})^{-2s} \sim (\log n)^{-\frac{2s}{p}}$.

3. **(super smooth - mildly ill-posed)** The variance term $m_k^2\frac{\sqrt{2k}}{n}$ is of order $\frac{1}{n}k^{1/2+2p}$ and the bias term a_k^2 is of order e^{-2k^s}. Hence, the optimal $k_{a_\bullet}^{\mathrm{d}}$ satisfies $e^{-2(k_{a_\bullet})^s} \sim \frac{1}{n}(k_{a_\bullet})^{1/2+2p}$ and thus $k_{a_\bullet}^{\mathrm{d}} \sim (\log(n/b_n))^{\frac{1}{s}}$ with $b_n \sim (\log n)^{\frac{4p+1}{2s}}$, which yields an upper bound of order $(\rho_{a_\bullet}^{\mathrm{d}})^2 \sim \frac{1}{n}(k_{a_\bullet}^{\mathrm{d}})^{2p+1/2} \sim \frac{1}{n}(\log n)^{\frac{2p+1/2}{s}}$.

□

4.5 Adaptive direct testing procedure

4.5.1 Aggregation of the direct tests and the choice of the levels α_k

The test $\Delta^{\mathrm{d}}_{k^{\mathrm{d}}_{a_\bullet},\alpha/2}$ in Corollary 4.4.3 requires the knowledge of the parameter sequence $a_\bullet$ of the regularity class $\mathcal{E}^{\mathrm{R}}_{a_\bullet}$ for the choice of the optimal dimension parameter $k = k^{\mathrm{d}}_{a_\bullet}$. Let $\mathcal{K} \subseteq \mathbb{N}$ be a finite collection of dimension parameters. We apply the Bonferroni aggregation method described in Section 4.3 to the collection of direct tests $(\Delta^{\mathrm{d}}_{k,\alpha_k})_{k\in\mathcal{K}}$ in order to construct an adaptive (i.e. *assumption-free*) testing procedure. We obtain a max-test with a Bonferroni choice of error levels $\alpha_k = \alpha/|\mathcal{K}|$

$$\Delta^{\mathrm{d}}_{\mathcal{K},\alpha} := \mathbb{1}_{\{\tilde{Q}_{\mathcal{K},\alpha}>0\}} \qquad \text{with} \qquad \tilde{Q}_{\mathcal{K},\alpha} := \max_{k\in\mathcal{K}} \left(\widetilde{\mathrm{q}}_k^2 - \tau_k^{\mathrm{d}}(\tfrac{\alpha}{|\mathcal{K}|})\right).$$

4.5.2 Testing radius of the direct max-test

In the next proposition we determine an adaptive upper bound for the radius of testing of the max-test. Again, as it is the case for the indirect-max-test, the upper bound has two regimes. The adaptive factor δ^{-1} depends on which regime governs the behaviour of the radius of testing. The adaptive factor δ^{-1} is in all cases of order $(\log|\mathcal{K}|)^c$ for $c \in \left\{\frac{1}{2}, 1\right\}$. Below, we give conditions for which the adaptive factor is of order $(\log|\mathcal{K}|)^{1/2}$, which we show to be the minimal adaptive factor. The max-test $\Delta^{\mathrm{d}}_{\mathcal{K},\alpha}$ only aggregates over a finite set $\mathcal{K} \subseteq \mathbb{N}$, therefore we define the minimal achievable radius of testing over the set $\mathcal{K}$ as

$$(\rho^{\mathrm{d}}_{\mathcal{K},a_\bullet}(n))^2 := \min_{k\in\mathcal{K}}(\rho^{\mathrm{d}}_{k,a_\bullet}(n))^2 \qquad \text{with} \qquad (\rho^{\mathrm{d}}_{k,a_\bullet}(n))^2 := a_k^2 \vee \frac{\sqrt{2k}}{n}m_k^2,$$

and m_k^2 as in (4.2.8) and a regularity parameter $a_\bullet = (a_j)_{j\in\mathbb{N}} \in \mathcal{A}$. Since $(\rho^{\mathrm{d}}_{a_\bullet}(n))^2$ in (4.4.9) is defined as the minimum taken over $\mathbb{N}$ instead of $\mathcal{K}$, for $n \in \mathbb{N}$ we always have $\rho^{\mathrm{d}}_{a_\bullet}(n) = \rho^{\mathrm{d}}_{\mathbb{N},a_\bullet}(n) \leq \rho^{\mathrm{d}}_{\mathcal{K},a_\bullet}(n)$. Moreover, let us recall the remainder radius defined in (4.3.4)

$$r^2_{\mathcal{K},a_\bullet}(n) = \min_{k\in\mathcal{K}} r^2_{k,a_\bullet}(n) \qquad \text{with} \qquad r^2_{k,a_\bullet}(n) = a_k^2 \vee \frac{m_k^2}{n}.$$

Proposition 4.5.1 (Uniform radius of testing over $\mathcal{A}$).
Under the assumptions of Proposition 4.4.1 let $\alpha \in (0,1)$ and consider $\overline{A}_\alpha$ as in (4.4.10). Then, for all $A \geq \overline{A}_\alpha$ and for all $n \in \mathbb{N}$, $n \geq 2$

$$\sup_{a_\bullet\in\mathcal{A}} \mathcal{R}\left(\Delta^{\mathrm{d}}_{\mathcal{K},\alpha/2} \mid \mathcal{E}^{\mathrm{R}}_{a_\bullet}, A\left(1 \vee \frac{\rho^{\mathrm{d}}_{\mathcal{K},a_\bullet}(\delta n)}{\delta^{3/2}}\right)\left(r_{\mathcal{K},a_\bullet}(\delta^2 n) \vee \rho^{\mathrm{d}}_{\mathcal{K},a_\bullet}(\delta n)\right)\right) \leq \alpha$$

with $\delta = (1 \vee \log|\mathcal{K}|)^{-1/2}$.

Proof of Proposition 4.5.1. The proof follows along the lines of the proof of Proposition 4.3.2 making use of Proposition 4.4.1 rather than Proposition 4.2.1. We again bound the type I and maximal type II error probabilities separately. From (4.3.1) combined with Proposition 4.4.1 (i) and $\sum_{k\in\mathcal{K}} \frac{\alpha}{2|\mathcal{K}|} = \frac{\alpha}{2}$ it follows that the type I error probability is bounded by $\alpha/2$. Under the alternative, let $f \in \mathscr{L}^2$ with $f - f^\circ \in \mathcal{E}^{\mathrm{R}}_{a_\bullet}$ satisfy

$$\|f - f^\circ\|^2_{\mathscr{L}^2} \geq \overline{A}^2_\alpha \left(1 \vee \frac{(\rho^{\mathrm{d}}_{\mathcal{K},a_\bullet}(\delta n))^2}{\delta^3}\right)\left(r^2_{\mathcal{K},a_\bullet}(\delta^2 n) \vee (\rho^{\mathrm{d}}_{\mathcal{K},a_\bullet}(\delta n))^2\right). \tag{4.5.1}$$

It is sufficient to use the elementary bound (4.3.2) together with the following two observations, which we show below.

1. Whenever $f \in \mathscr{L}^2$ satisfies $f - f^\circ \in \mathcal{E}^{\mathrm{R}}_{a_\bullet}$ and

$$\|f - f^\circ\|^2_{\mathscr{L}^2} \geq \overline{A}^2_\alpha \left(a_k^2 \vee \left(1 \vee \frac{(2k)^{1/4}}{\delta^2 n^{1/2}} \vee \frac{\sqrt{2k}}{\delta^3 n}\right)\frac{\sqrt{2k} m_k^2}{\delta n} \vee \frac{m_k^2}{\delta^2 n}\right), \tag{4.5.2}$$

   then

$$\mathbb{P}_f\left(\widetilde{\mathrm{q}}^2_k < \tau^{\mathrm{d}}_k\left(\tfrac{\alpha}{2|\mathcal{K}|}\right)\right) \leq \frac{\alpha}{2}.$$

2. If the separation condition (4.5.1) is satisfied, then there exists a $k \in \mathcal{K}$ such that (4.5.2) is fulfilled.

Consequently, we have

$$\mathbb{P}_f(\Delta_{\mathcal{K},\alpha/2} = 0) \leq \min_{k\in\mathcal{K}} \mathbb{P}_f(\Delta_{k,\alpha_k/2} = 0) \leq \frac{\alpha}{2}.$$

for all $f \in \mathscr{L}^2$ satisfying $f - f^\circ \in \mathcal{E}^{\mathrm{R}}_{a_\bullet}$ and (4.3.5). Thus, the maximal type II error probability is also bounded by $\alpha/2$. It remains to show (1.) and (2.).

1. The claim follows from Proposition 4.4.1 (ii) (with $\beta = \alpha/2$), since (4.5.2) implies (4.4.6), which states

$$\mathrm{q}^2_k(g - g^\circ) \geq 2\left(\tau^{\mathrm{d}}_k(\tfrac{\alpha}{2|\mathcal{K}|}) + c_3 L^4_{\alpha/4}\left(1 \vee \frac{\sqrt{2k}}{n}\right)\frac{\sqrt{2k}}{n}\right), \tag{4.5.3}$$

   with $\tau_k(\frac{\alpha}{2|\mathcal{K}|})$ as in (4.2.9). Indeed, exploiting $L^2_{\alpha/2} = \log(2e/\alpha) \geq 1$ and, hence,

$$L^2_{\alpha/(2|\mathcal{K}|)} = \log(2e\,|\mathcal{K}|\,/\alpha) = \log|\mathcal{K}| + L^2_{\alpha/2} \leq L^2_{\alpha/2}(1 + \log|\mathcal{K}|) = L^2_{\alpha/2}\delta^{-2},$$

   we have

$$\begin{aligned}
\tau^{\mathrm{d}}_k(\tfrac{\alpha}{2|\mathcal{K}|}) &= c_1\left(1 \vee L^2_{\alpha/(2|\mathcal{K}|)}\sqrt{\frac{\sqrt{2k}}{n}} \vee L^3_{\alpha/(2|\mathcal{K}|)}\frac{\sqrt{2k}}{n}\right) L_{\alpha/(2|\mathcal{K}|)}\frac{\sqrt{2k}}{n} + c_2 L^2_{\alpha/(2|\mathcal{K}|)}\frac{1}{n} \\
&= c_1\left(1 \vee L^2_{\alpha/2}\sqrt{\frac{\sqrt{2k}}{\delta^4 n}} \vee L^3_{\alpha/2}\frac{\sqrt{2k}}{\delta^3 n}\right) L_{\alpha/2}\frac{\sqrt{2k}}{\delta n} + c_2 L^2_{\alpha/2}\frac{1}{\delta^2 n} \\
&\leq (c_1 + c_2) L^4_{\alpha/2}\left(\left(1 \vee \sqrt{\frac{\sqrt{2k}}{\delta^4 n}} \vee \frac{\sqrt{2k}}{\delta^3 n}\right)\frac{\sqrt{2k}}{\delta n} \vee \frac{1}{\delta^2 n}\right).
\end{aligned}$$

Additionally using $L_{\alpha/4} \geq L_{\alpha/2}$, $1 \geq \delta$ the right-hand side of (4.5.3) is bounded by

$$
\begin{aligned}
&2\left(\tau_k^{\mathrm{d}}(\tfrac{\alpha}{2|\mathcal{K}|}) + c_3 L_{\alpha/4}^4 \left(1 \vee \frac{\sqrt{2k}}{n}\right)\frac{\sqrt{2k}}{n}\right) \\
&\leq 2(c_1+c_2)L_{\alpha/2}^4 \left(\left(1 \vee \sqrt{\frac{\sqrt{2k}}{\delta^4 n}} \vee \frac{\sqrt{2k}}{\delta^3 n}\right)\frac{\sqrt{2k}}{\delta n} \vee \frac{1}{\delta^2 n}\right) + 2c_3 L_{\alpha/4}^4 \left(1 \vee \frac{\sqrt{2k}}{n}\right)\frac{\sqrt{2k}}{n} \\
&\leq 2(c_1+c_2+c_3)L_{\alpha/4}^4 \left(\left(1 \vee \sqrt{\frac{\sqrt{2k}}{\delta^4 n}} \vee \frac{\sqrt{2k}}{\delta^3 n}\right)\frac{\sqrt{2k}}{\delta n} \vee \frac{1}{\delta^2 n}\right).
\end{aligned}
$$

Hence, since $\overline{A}_\alpha^2 - R^2 \geq 2(c_1+c_2+c_3)L_{\alpha/4}^4$ the condition (4.4.6) of Proposition 4.4.1 holds whenever

$$
\mathrm{q}_k^2(f - f^\circ) \geq \left(\overline{A}_\alpha^2 - R^2\right)\left(\left(1 \vee \sqrt{\frac{\sqrt{2k}}{\delta^4 n}} \vee \frac{\sqrt{2k}}{\delta^3 n}\right)\frac{\sqrt{2k}}{\delta n} \vee \frac{1}{\delta^2 n}\right). \tag{4.5.4}
$$

Let us verify this condition. Due to $f - f^\circ \in \mathcal{E}_{a_\bullet}^{\mathrm{R}}$ and, hence, $\sum_{|j|>k}\left|f_j^\circ - f_j\right|^2 \leq a_k^2 \mathrm{R}^2$, the condition (4.5.2) implies

$$
\begin{aligned}
\mathrm{q}_k^2(f - f^\circ) &= \|f - f^\circ\|_{\mathscr{L}^2}^2 - \sum_{|j|>k}\left|f_j^\circ - f_j\right|^2 \\
&\geq \left(\overline{A}_\alpha^2 - R^2\right)\left(\left(1 \vee \sqrt{\frac{\sqrt{2k}}{\delta^4 n}} \vee \frac{\sqrt{2k}}{\delta^3 n}\right)\frac{\sqrt{2k}}{\delta n} m_k^2 \vee \frac{m_k^2}{\delta^2 n}\right),
\end{aligned}
$$

and, hence, since $m_k^2 \mathrm{q}_k^2(g - g^\circ) \geq \mathrm{q}_k^2(f - f^\circ)$, we obtain

$$
\mathrm{q}_k^2(g - g^\circ) \geq \left(\overline{A}_\alpha^2 - R^2\right)\left(\left(1 \vee \sqrt{\frac{\sqrt{2k}}{\delta^4 n}} \vee \frac{\sqrt{2k}}{\delta^3 n}\right)\frac{\sqrt{2k}}{\delta n} \vee \frac{1}{\delta^2 n}\right),
$$

which justifies the application of Proposition 4.4.1. If (4.5.2) is satisfied, then also (4.5.4) and thus (4.4.6), which shows the claim (1.).

2. By the Balancing Lemma A.2.1 we have

$$
r_{\mathcal{K},a_\bullet}^2(\delta^2 n) \vee (\rho_{\mathcal{K},a_\bullet}^{\mathrm{d}}(\delta n))^2 = a_k^2 \vee \frac{m_k^2}{n} \vee \frac{\sqrt{2k}}{\delta n} m_k^2
$$

$$
\text{and} \qquad (\rho_{\mathcal{K},a_\bullet}^{\mathrm{d}}(\delta n))^2 \geq \frac{\sqrt{2k}}{\delta n} m_k^2 \geq \frac{\sqrt{2k}}{\delta n}
$$

for at least one $k \in \mathcal{K}$. Hence, there exists a dimension parameter $k \in \mathcal{K}$ such that

$$
\begin{aligned}
&\left(r_{\mathcal{K},a_\bullet}^2(\delta^2 n) \vee (\rho_{\mathcal{K},a_\bullet}^{\mathrm{d}}(\delta n))^2\right)\left(1 \vee \frac{(\rho_{\mathcal{K},a_\bullet}^{\mathrm{d}}(\delta n))^2}{\delta^3}\right) \\
&\qquad \geq a_k^2 \vee \frac{m_k^2}{\delta^2 n} \vee \left(\frac{\sqrt{2k}}{\delta n} m_k^2 \left(1 \vee \frac{\sqrt{2k}}{\delta^4 n}\right)\right).
\end{aligned}
$$

Since $1 \vee \frac{\sqrt{2k}}{\delta^4 n} \geq 1 \vee \frac{(2k)^{1/4}}{\delta^2 n^{1/2}} \vee \frac{\sqrt{2k}}{\delta^3 n}$, this shows (4.5.2) and, hence, (2.), which completes the proof.

□

The next two corollaries show that we either (in the worst case) obtain an adaptive factor of order $\log|\mathcal{K}|$ or (in the best case) of order $(\log|\mathcal{K}|)^{1/2}$, depending on whether the remainder term $r_{\mathcal{K},a_\bullet}(\delta^2 n)$ is negligible compared with $\rho_{\mathcal{K},a_\bullet}^{\mathrm{d}}(\delta n)$ or not.

Corollary 4.5.2 (Worst-case adaptive factor).
Under the assumptions of Proposition 4.4.1, let $\alpha \in (0,1)$ and consider $\overline{A}_\alpha$ as in (4.4.10). Then, for all $A \geq \overline{A}_\alpha$ and for all $n \in \mathbb{N}$, $n \geq 2$,

$$\sup_{a_\bullet \in \mathcal{A}} \mathcal{R}\left(\Delta^{\mathrm{d}}_{\mathcal{K},\alpha/2} \mid \mathcal{E}^{\mathrm{R}}_{a_\bullet}, A\left(1 \vee \rho^{\mathrm{d}}_{\mathcal{K},a_\bullet}(\delta^2 n)\right)\rho^{\mathrm{d}}_{\mathcal{K},a_\bullet}(\delta^2 n)\right) \leq \alpha$$

with $\delta = (1 \vee \log|\mathcal{K}|)^{-1/2}$.

Proof of Corollary 4.5.2. The proof follows along the lines of the proof of Proposition 4.3.2, considering

$$\|f - f^\circ\|^2_{\mathscr{L}^2} \geq \overline{A}^2_\alpha \left(1 \vee (\rho^{\mathrm{d}}_{\mathcal{K},a_\bullet}(\delta^2 n))^2\right)(\rho^{\mathrm{d}}_{\mathcal{K},a_\bullet}(\delta^2 n))^2 \tag{4.5.5}$$

instead of (4.5.1). In fact, as in the proof of Proposition 4.5.1, it is sufficient to show (2.) under the separation condition (4.5.5). For each $a_\bullet \in \mathcal{A}$ under (4.5.5) the dimension parameter $k_\star := \arg\min_{k\in\mathcal{K}} \rho^{\mathrm{d}}_{k,a_\bullet}(\delta^2 n)$ satisfies

$$\begin{aligned}
&\left(1 \vee (\rho^{\mathrm{d}}_{\mathcal{K},a_\bullet}(\delta^2 n))^2\right)(\rho^{\mathrm{d}}_{\mathcal{K},a_\bullet}(\delta^2 n))^2 \\
&= \left(1 \vee \rho^{\mathrm{d}}_{\mathcal{K},a_\bullet}(\delta^2 n) \vee (\rho^{\mathrm{d}}_{\mathcal{K},a_\bullet}(\delta^2 n))^2\right)(\rho^{\mathrm{d}}_{\mathcal{K},a_\bullet}(\delta^2 n))^2 \\
&\geq a^2_{k_\star} \vee \frac{\sqrt{2k_\star}}{\delta^2 n} m^2_{k_\star}\left(1 \vee \frac{(2k_\star)^{1/4}}{\delta n^{1/2}} m_{k_\star} \vee \frac{\sqrt{2k_\star}}{\delta^2 n} m^2_{k_\star}\right) \\
&\geq a^2_{k_\star} \vee \frac{\sqrt{2k_\star}}{\delta^2 n} m^2_{k_\star}\left(1 \vee \frac{(2k_\star)^{1/4}}{\delta n^{1/2}} m_{k_\star} \vee \frac{\sqrt{2k_\star}}{\delta^2 n} m^2_{k_\star}\right) \vee \frac{m^2_{k_\star}}{\delta^2 n} \\
&\geq a^2_{k_\star} \vee \frac{\sqrt{2k_\star}}{\delta n} m^2_{k_\star}\left(\frac{1}{\delta} \vee \frac{(2k_\star)^{1/4}}{\delta^2 n^{1/2}} m_{k_\star} \vee \frac{\sqrt{2k_\star}}{\delta^3 n} m^2_{k_\star}\right) \vee \frac{m^2_{k_\star}}{\delta^2 n} \\
&\geq a^2_{k_\star} \vee \frac{\sqrt{2k_\star}}{\delta n} m^2_{k_\star}\left(1 \vee \frac{(2k_\star)^{1/4}}{\delta^2 n^{1/2}} m_{k_\star} \vee \frac{\sqrt{2k_\star}}{\delta^3 n} m^2_{k_\star}\right) \vee \frac{m^2_{k_\star}}{\delta^2 n},
\end{aligned}$$

since $a^2_{k_\star} \vee \frac{\sqrt{2k_\star} m^2_{k_\star}}{\delta^2 n} = (\rho^{\mathrm{d}}_{\mathcal{K},a_\bullet}(\delta^2 n))^2$ and $\delta \leq 1$. This shows (4.5.2) and, consequently, (2.). We obtain the assertion by proceeding exactly as in the proof of Proposition 4.5.1. □

Corollary 4.5.2 implies that $(\rho^{\mathrm{d}}_{\mathcal{K},a_\bullet}(\delta^2 n))^2$ is an upper bound for the radius of testing for the direct max-test if $\rho^{\mathrm{d}}_{\mathcal{K},a_\bullet}(\delta^2 n) \leq 1$. Note that this is the case for an arbitrary regularity parameter $a_\bullet \in \mathcal{A}$, if e.g. $1 \in \mathcal{K}$ and $n \geq \sqrt{2}\,|\varphi_1|^{-2}(1 + \log|\mathcal{K}|)$, that is, for a suitable choice of $\mathcal{K}$ and n large enough. Indeed, under these two conditions, we have

$$(\rho^{\mathrm{d}}_{\mathcal{K},a_\bullet}(\delta^2 n))^2 = \arg\min_{k\in\mathcal{K}}\left\{a^2_k \vee \frac{\sqrt{2k} m^2_k}{\delta^2 n}\right\} \leq a^2_1 \vee \frac{\sqrt{2} m^2_1}{\delta^2 n} \leq 1 \vee \frac{\sqrt{2}\,|\varphi_1|^{-2}}{\delta^2 n} = 1.$$

Under additional conditions, which are satisfied for all examples considered in our illustrations, we can derive a sharper upper bound $(\rho^{\mathrm{d}}_{\mathcal{K},a_\bullet}(\delta n))^2$ and, thus, obtain an adaptive factor of order $(\log|\mathcal{K}|)^{1/2}$.

Corollary 4.5.3 (Best-case adaptive factor). Under the assumptions of Proposition 4.4.1, let $\alpha \in (0,1)$ and consider $\overline{A}_\alpha$ as in (4.4.10). If there exist constants $c, C > 1$ such that

$$r_{\mathcal{K},a_\bullet}(\delta^2 n) \leq c\rho^{\mathrm{d}}_{\mathcal{K},a_\bullet}(\delta n) \qquad \text{and} \qquad \rho^{\mathrm{d}}_{\mathcal{K},a_\bullet}(\delta n) \leq C\delta^{3/2} \tag{4.5.6}$$

for all $a_\bullet \in \mathcal{A}$, then for all $A \geq c \cdot C \cdot \overline{A}_\alpha$ and for all $n \in \mathbb{N}$, $n \geq 2$,

$$\sup_{a_\bullet \in \mathcal{A}} \mathcal{R}\left(\Delta^{\mathrm{d}}_{\mathcal{K},\alpha/2} \mid \mathcal{E}^{\mathrm{R}}_{a_\bullet}, A\rho^{\mathrm{d}}_{\mathcal{K},a_\bullet}(\delta n)\right) \leq \alpha$$

with $\delta = (1 \vee \log|\mathcal{K}|)^{-1/2}$.

Proof of Corollary 4.5.3. Under the assumptions (4.5.6) we have

$$\begin{aligned}\left(1 \vee \frac{\rho^{\mathrm{d}}_{\mathcal{K},a_\bullet}(\delta n)}{\delta^{3/2}}\right)\left(r_{\mathcal{K},a_\bullet}(\delta^2 n) \vee \rho^{\mathrm{d}}_{\mathcal{K},a_\bullet}(\delta n)\right) &\leq \left(1 \vee \frac{\rho^{\mathrm{d}}_{\mathcal{K},a_\bullet}(\delta n)}{\delta^{3/2}}\right)\left(c\rho^{\mathrm{d}}_{\mathcal{K},a_\bullet}(\delta n) \vee \rho^{\mathrm{d}}_{\mathcal{K},a_\bullet}(\delta n)\right)\\ &\leq (1 \vee C)\left(c\rho^{\mathrm{d}}_{\mathcal{K},a_\bullet}(\delta n) \vee \rho^{\mathrm{d}}_{\mathcal{K},a_\bullet}(\delta n)\right)\\ &\leq c \cdot C \cdot \rho^{\mathrm{d}}_{\mathcal{K},a_\bullet}(\delta n),\end{aligned}$$

hence, the assertion follows directly from Proposition 4.5.1. □

Concerning the choice of the collection $\mathcal{K}$ of dimensions, we refer to Remark 4.3.5.

Illustration 4.5.4. For the typical configurations for regularity and ill-posedness introduced in Illustration 4.2.5 the tables below display the adaptive radii of the direct max-test $\Delta^{\mathrm{d}}_{\mathcal{K},\alpha/2}$, $\alpha \in (0,1)$ for appropriately chosen grids. In a mildly ill-posed model (parameter p) with ordinary smoothness (parameter s) we have seen in Illustration 4.4.5 that the optimal dimension $k_{a_\bullet}$ is of order $n^{\frac{2}{4p+4s+1}}$, which is smaller than n^2 for all combinations of s and p, by the reasoning of Remark 4.3.5 it is even smaller than $n^2/2$. Hence, we choose the geometric grid

$$\mathcal{K}_2 = \left\{2^j, j \in \left\{0,\dots,\lfloor\log_2(\tfrac{n^2}{2})\rfloor\right\}\right\} \qquad \text{with} \qquad |\mathcal{K}_2| = \lfloor\log_2(\tfrac{n^2}{2})\rfloor$$

and obtain the adaptive factor $\delta^{-1} = (1+\log|\mathcal{K}_2|)^{1/2} \sim \sqrt{\log\log n}$. It is easily seen that the remainder term $r^2_{\mathcal{K}_2,a_\bullet}(\delta^2 n)$ is asymptotically negligible compared with $(\rho^{\mathrm{d}}_{\mathcal{K}_2,a_\bullet}(\delta n))^2$, since for some positive constants $x, y > 0$ (depending on s and p) we have

$$\frac{r^2_{\mathcal{K}_2,a_\bullet}(\delta^2 n)}{(\rho^{\mathrm{d}}_{\mathcal{K}_2,a_\bullet}(\delta n))^2} \sim \frac{(\log\log n)^x}{n^y} \longrightarrow 0 \qquad (n \to \infty).$$

Moreover, $\frac{(\rho^{\mathrm{d}}_{\mathcal{K}_2,a_\bullet}(\delta n))^2}{\delta^3}$ tends to zero for $n \to \infty$, since δ is only of log-order. Therefore, the upper bound derived in Proposition 4.5.1 reduces to $(\rho^{\mathrm{d}}_{\mathcal{K}_2,a_\bullet}(\delta n))^2$, which is of the same order as $(\rho^{\mathrm{d}}_{a_\bullet}(\delta n))^2$ with an adaptive factor of order $\sqrt{\log\log n}$.
In a severely ill-posed model with ordinary smoothness, we have seen in Illustration 4.4.5 that the order of the optimal dimension parameter does not depend on the smoothness parameter $a_\bullet$. Hence, the test $\Delta^{\mathrm{d}}_{\alpha/2,k_{a_\bullet}}$, which is, in fact, independent of $a_\bullet$, is automatically adaptive with respect to ordinary smoothness and no aggregation procedure is needed. Note that in the case $f^\circ = \mathbb{1}_{[0,1)}$ neither the test statistic $\widetilde{\mathrm{q}}^2_k$ (defined in (4.4.2)) nor the threshold $\tau^{\mathrm{d}}_k(\alpha)$ (defined in (4.4.5)) depend on the coefficients of the error density φ. Then, by aggregating over $\mathcal{K}_2$ the max-test $\Delta^{\mathrm{d}}_{\mathcal{K},\alpha/2}$ is also adaptive with respect to severe ill-posedness. In this case the remainder term $r^2_{\mathcal{K}_2,a_\bullet}(\delta^2 n)$ is of the same order as $(\rho^{\mathrm{d}}_{\mathcal{K}_2,a_\bullet}(\delta n))^2$, hence the theoretical adaptive factor that we obtain is of order $\log\log n$. This factor, however, does not effect the behaviour of the radius. Moreover, $\frac{(\rho^{\mathrm{d}}_{\mathcal{K}_2,a_\bullet}(\delta n))^2}{\delta^3}$ tends to zero for $n \to \infty$, since δ is only of $\log\log$-order and, therefore, the upper bound in Proposition 4.5.1 asymptotically reduces to $(\rho^{\mathrm{d}}_{\mathcal{K}_2,a_\bullet}(\delta n))^2$, which is of the same order as $(\rho^{\mathrm{d}}_{a_\bullet}(n))^2$.

Order of $r^2_{\mathcal{K}_2,a_\bullet}(\delta^2 n)$ and $(\rho^{\mathrm{d}}_{\mathcal{K}_2,a_\bullet}(\delta n))^2$ with $\delta = (1+\log|\mathcal{K}_2|)^{-1/2}$ and $\mathcal{K}_2 = \left\{2^j, j \in \left\{0,\dots,\lfloor \log_2(\frac{n^2}{2})\rfloor\right\}\right\}$			
a_j (smoothness)	$\lvert\varphi_j\rvert$ (ill-posedness)	$r^2_{\mathcal{K}_2,a_\bullet}(\delta^2 n)$	$(\rho^{\mathrm{d}}_{\mathcal{K}_2,a_\bullet}(\delta n))^2$
j^{-s}	$\lvert j\rvert^{-p}$	$\left(\frac{n}{\log\log n}\right)^{-\frac{4s}{4s+4p}}$	$\left(\frac{n}{(\log\log n)^{1/2}}\right)^{-\frac{4s}{4s+4p+1}}$
j^{-s}	$e^{-\lvert j\rvert^p}$	$(\log n)^{\frac{1}{p}}$	$(\log n)^{-\frac{2s}{p}}$

In a mildly ill-posed model with super smoothness (parameter s) Illustration 4.4.5 shows that the optimal dimension is of order $(\log n)^{\frac{1}{s}}$. Hence, we choose the smaller geometric grid

$$\mathcal{K}_{s_\star} = \left\{2^j, j \in \left\{0,\dots,\lfloor \tfrac{\log_2 \log n}{s_\star}\rfloor\right\}\right\} \qquad \text{with} \qquad |\mathcal{K}_{s_\star}| = \lfloor \tfrac{\log_2 \log n}{s}\rfloor$$

for adaptation to $s \geq s_\star$ and obtain the adaptive factor $\delta^{-1} = (1+\log|\mathcal{K}_{s_\star}|)^{1/2}$. It is easily seen that the remainder term $r^2_{\mathcal{K}_{s_\star},a_\bullet}(\delta^2 n)$ is asymptotically negligible compared with $(\rho^{\mathrm{d}}_{\mathcal{K}_{s_\star},a_\bullet}(\delta n))^2$, since

$$\frac{r^2_{\mathcal{K}_{s_\star},a_\bullet}(\delta^2 n)}{(\rho_{\mathcal{K}_{s_\star},a_\bullet}(\delta n))^2} \sim \frac{\sqrt{\log\log\log n}}{(\log n)^{\frac{1}{2s}}} \longrightarrow 0 \qquad (n \to \infty).$$

Moreover, $\frac{(\rho^{\mathrm{d}}_{\mathcal{K}_{s_\star},a_\bullet}(\delta n))^2}{\delta^3}$ tends to zero for $n \to \infty$, since δ is only of log-order. Therefore, the upper bound derived in Proposition 4.3.2 asymptotically reduces to $(\rho^{\mathrm{d}}_{\mathcal{K}_{s_\star},a_\bullet}(\delta n))^2$, which is of the same order as $(\rho^{\mathrm{d}}_{a_\bullet}(\delta n))^2$ with an adaptive factor of order $\sqrt{\log\log\log n}$.

Order of $r^2_{\mathcal{K}_{s_\star},a_\bullet}(\delta^2 n)$ and $(\rho^{\mathrm{d}}_{\mathcal{K}_{s_\star},a_\bullet}(\delta n))^2$ with $\delta = (1+\log|\mathcal{K}_{s_\star}|)^{-1/2}$ and $\mathcal{K}_{s_\star} = \left\{2^j, j \in \left\{0,\dots,\lfloor \frac{\log_2 \log n}{s_\star}\rfloor\right\}\right\}$			
a_j (smoothness)	$\lvert\varphi_j\rvert$ (ill-posedness)	$r^2_{\mathcal{K}_{s_\star},a_\bullet}(\delta^2 n)$	$(\rho^{\mathrm{d}}_{\mathcal{K}_{s_\star},a_\bullet}(\delta n))^2$
e^{-j^s}	$\lvert j\rvert^{-p}$	$\frac{\log\log\log n}{n}(\log n)^{\frac{2p}{s}}$	$\frac{(\log\log\log n)^{1/2}}{n}(\log n)^{\frac{2p+1/2}{s}}$

We conclude that in all the cases considered in this illustration the direct max-test achieves a testing radius of the same order as the indirect max-test. We emphasise that in contrast to the indirect max-tests the direct max-tests are in addition also adaptive to the ill-posedness of the model, since they do not require the knowledge of the coefficients of the error density φ if we test for uniformity (i.e. in the case $f^\circ = \mathbb{1}_{[0,1)}$).

Calculations for the risk bounds in Illustration 4.5.4.
Firstly, we determine the order of the terms $(\rho^{\mathrm{d}}_{\mathcal{K},a_\bullet}(\delta n))^2 = \min_{k\in\mathcal{K}}\left\{a_k^2 \vee \frac{\sqrt{2k}}{\delta n} m_k^2\right\}$.

1. **(ordinary smooth - mildly ill-posed)**
   We first show that minimisation over $\mathcal{K}_2$ approximates the minimisation over $\mathbb{N}$ well enough, i.e.

$$(\rho^{\mathrm{d}}_{\mathcal{K}_2,a_\bullet}(n))^2 = \min_{k\in\mathcal{K}_2}(\rho^{\mathrm{d}}_{k,a_\bullet}(n))^2 \sim \min_{k\in\mathbb{N}}(\rho^{\mathrm{d}}_{a_\bullet}(n))^2 = (\rho^{\mathrm{d}}_{a_\bullet}(n))^2.$$

We aim to find $j_\star$ such that $2^{j_\star} \in \mathcal{K}_2$ approximates $k^{\mathrm{d}}_{a_\bullet}$ well. Since $k^{\mathrm{d}}_{a_\bullet} \sim n^{\frac{2}{4p+4s+1}}$ we define

$$j_\star := \lceil \tfrac{2}{4p+4s+1} \log_2 n \rceil \lesssim \lceil \log(n^2/2) \rceil.$$

Then,

$$\begin{aligned}(\rho^{\mathrm{d}}_{\mathcal{K}_2,a_\bullet}(n))^2 \leq (\rho^{\mathrm{d}}_{2^{j_\star},a_\bullet}(n))^2 = a^2_{2^{j_\star}} \vee \frac{\sqrt{2\cdot 2^{j_\star}} m^2_{2^{j_\star}}}{n} &\lesssim 2^{-2sj_\star} \vee \frac{2^{j_\star(2p+1/2)}}{n} \\ &\lesssim 2^{-(\frac{4s}{4p+4s+1}\log_2 n)} \vee \frac{2^{(\frac{2}{4p+4s+1}\log_2 n+1)(2p+1/2)}}{n} \\ &\lesssim n^{-\frac{4s}{4p+4s+1}} \vee 2^{2p+1/2} n^{-\frac{4s}{4p+4s+1}} \\ &\lesssim n^{-\frac{4s}{4p+4s+1}} \sim (\rho^{\mathrm{d}}_{a_\bullet}(n))^2\end{aligned}$$

Since, trivially $(\rho^{\mathrm{d}}_{a_\bullet}(n)) \leq (\rho^{\mathrm{d}}_{\mathcal{K}_2,a_\bullet}(n))^2$, we obtain $(\rho^{\mathrm{d}}_{a_\bullet}(n))^2 \sim (\rho^{\mathrm{d}}_{\mathcal{K}_2,a_\bullet}(n))^2$. Replacing n by δn yields the result.

2. **(super smooth - mildly ill-posed)**
We first show that minimisation over $\mathcal{K}_{s_\star}$ approximates the minimisation over $\mathbb{N}$ well enough, i.e.

$$(\rho^{\mathrm{d}}_{\mathcal{K}_{s_\star},a_\bullet}(n))^2 = \min_{k\in\mathcal{K}_{s_\star}} (\rho^{\mathrm{d}}_{k,a_\bullet}(n))^2 \sim \min_{k\in\mathbb{N}} (\rho^{\mathrm{d}}_{a_\bullet}(n))^2 = (\rho_{a_\bullet}(n))^2.$$

We aim to find $j_\star$ such that $2^{j_\star} \in \mathcal{K}_{s_\star}$ approximates $k^{\mathrm{d}}_{a_\bullet}$ well. Since $k^{\mathrm{d}}_{a_\bullet} \sim (\log n)^{\frac{1}{s}}$ we define

$$j_\star := \lceil \tfrac{1}{s} \log_2 \log n \rceil \lesssim \lceil \tfrac{1}{s_\star} \log\log(n) \rceil.$$

Then,

$$\begin{aligned}(\rho^{\mathrm{d}}_{\mathcal{K}_{s_\star},a_\bullet}(n))^2 \leq (\rho^{\mathrm{d}}_{2^{j_\star},a_\bullet}(n))^2 = a^2_{2^{j_\star}} \vee \frac{\sqrt{2\cdot 2^{j_\star}} m^2_{2^{j_\star}}}{n} &\lesssim e^{-2\cdot 2^{sj_\star}} \vee \frac{2^{j_\star(2p+1/2)}}{n} \\ &\lesssim e^{-2\cdot 2^{s\frac{1}{s}\log_2\log n}} \vee \frac{2^{(\frac{1}{s}\log_2\log n+1)(2p+1/2)}}{n} \\ &\lesssim n^{-2} \vee 2^{2p+1/2} \frac{(\log n)^{\frac{2p+1/2}{s}}}{n} \\ &\lesssim \frac{(\log n)^{\frac{2p+1/2}{s}}}{n} \sim (\rho^{\mathrm{d}}_{a_\bullet}(n))^2\end{aligned}$$

Since, trivially $(\rho^{\mathrm{d}}_{a_\bullet}(n)) \leq (\rho^{\mathrm{d}}_{\mathcal{K}_{s_\star},a_\bullet}(n))^2$, we obtain $(\rho^{\mathrm{d}}_{a_\bullet}(n))^2 \sim (\rho^{\mathrm{d}}_{\mathcal{K}_{s_\star},a_\bullet}(n))^2$. Replacing n by δn yields the result.

3. **(ordinary smooth - severely ill-posed)** We first show that minimisation over $\mathcal{K}_2$ approximates the minimisation over $\mathbb{N}$ well enough, i.e.

$$(\rho^{\mathrm{d}}_{\mathcal{K}_2,a_\bullet}(n))^2 = \min_{k\in\mathcal{K}_2} (\rho^{\mathrm{d}}_{k,a_\bullet}(n))^2 \sim \min_{k\in\mathbb{N}} (\rho^{\mathrm{d}}_{a_\bullet}(n))^2 = (\rho^{\mathrm{d}}_{a_\bullet}(n))^2.$$

We aim to find $j_\star$ such that $2^{j_\star} \in \mathcal{K}_2$ approximates $k^{\mathrm{d}}_{a_\bullet}$ well. Since $k^{\mathrm{d}}_{a_\bullet} \sim (\log n/b_n)^{\frac{1}{p}}$ with $b_n \sim (\log n)^{\frac{4s+1}{2p}}$ we define

$$j_\star := \lfloor \tfrac{1}{p} \log_2(\log(n/b_n)/2) \rfloor \lesssim \lceil \log(n^2/2) \rceil.$$

Then,

$$(\rho^{\mathrm{d}}_{\mathcal{K}_2,a_\bullet}(n))^2 \leq (\rho^{\mathrm{d}}_{2^{j_\star},a_\bullet}(n))^2 = a^2_{2^{j_\star}} \vee \frac{\sqrt{2\cdot 2^{j_\star}}m^2_{2^{j_\star}}}{n} \lesssim 2^{-2sj_\star} \vee \frac{2^{j_\star/2}}{n}e^{2\cdot 2^{pj_\star}}$$
$$\lesssim 2^{-2s(\frac{1}{p}\log_2(\log(n/b_n)/2)-1)} \vee \frac{2^{\frac{1}{p}\log_2\log n}}{n}e^{2\cdot 2^{\log_2((\log(n/b_n))/2)}}$$
$$\lesssim (\log n)^{-\frac{2s}{p}} \sim (\rho^{\mathrm{d}}_{a_\bullet}(n))^2$$

Since, trivially $(\rho^{\mathrm{d}}_{a_\bullet}(n)) \leq (\rho^{\mathrm{d}}_{\mathcal{K}_2,a_\bullet}(n))^2$, we obtain $(\rho^{\mathrm{d}}_{a_\bullet}(n))^2 \sim (\rho^{\mathrm{d}}_{\mathcal{K}_2,a_\bullet}(n))^2$. Replacing n by δn yields the result.

Next, we determine the order of the remainder term $r^2_{\mathcal{K},a_\bullet}(\delta n) = \min_{k\in\mathcal{K}}\left\{a_k^2 \vee \frac{m_k^2}{n}\right\}$, by first calculating $r^2_{\mathbb{N},a_\bullet}(n) = \min_{k\in\mathbb{N}}\left\{a_k^2 \vee \frac{m_k^2}{n}\right\}$ and then showing that minimisation over $\mathcal{K}$ approximates the minimisation over $\mathbb{N}$ well enough. The calculations in the **(ordinary smooth - mildly ill-posed)** and the **(super smooth - mildly ill-posed)** cases have already been done in Illustration 4.3.6. It remains to consider the third case.

3. **(ordinary smooth - severely ill-posed)** The variance term $\frac{m_k^2}{n}$ is of order $\frac{e^{2k^p}}{n}$ and the bias term a_k^2 is of order k^{-2s}. Hence, the minimizing $k_\star$ satisfies $k_\star^{-2s} \sim \frac{e^{2k_\star^p}}{n}$ and thus $k_\star \sim (\log(n/b_n))^{\frac{1}{p}}$ with $b_n \sim (\log n)^{-\frac{2s}{p}}$, which yields $r^2_{\mathbb{N},a_\bullet} \sim (\log n)^{-\frac{2s}{p}}$.
   Next, we show that minimisation over $\mathcal{K}_2$ approximates the minimization over $\mathbb{N}$ well enough, i.e.

   $$r^2_{\mathcal{K}_2,a_\bullet}(n) = \min_{k\in\mathcal{K}_2} r^2_{k,a_\bullet}(n) \sim \min_{k\in\mathbb{N}} r^2_{k,a_\bullet}(n) = r^2_{a_\bullet}(n)$$

   We aim to find $j_\star$ such that $2^{j_\star} \in \mathcal{K}_2$ approximates $k_\star$ well. Since $k_\star \sim (\log(n/b_n))^{\frac{1}{p}}$ with $b_n \sim (\log n)^{-\frac{2s}{p}}$ we define

   $$j_\star := \lfloor \tfrac{1}{p}\log_2(\log(n/b_n)/2)\rfloor \lesssim \lfloor \log(n^2/2)\rfloor.$$

   Then,

   $$r^2_{\mathcal{K}_2,a_\bullet}(n) \leq r^2_{2^{j_\star},a_\bullet} = a^2_{2^{j_\star}} \vee \frac{m^2_{2^{j_\star}}}{n} \lesssim 2^{-2sj_\star} \vee \frac{e^{2\cdot 2^{pj_\star}}}{n}$$
   $$\lesssim 2^{-\frac{2s}{p}\log_2\log(n)} \vee \frac{e^{2\cdot 2^{\log_2(\log(n/b_n)/2)}}}{n}$$
   $$\lesssim (\log n)^{-\frac{2s}{p}} \sim r^2_{\mathbb{N},a_\bullet}(n)$$

   Since, trivially $r^2_{\mathbb{N},a_\bullet}(n) \leq r^2_{\mathcal{K}_2,a_\bullet}(n)$, we obtain $r^2_{\mathbb{N},a_\bullet}(n) \sim r^2_{\mathcal{K}_2,a_\bullet}(n)$. Replacing n by $\delta^2 n$ yields the result.

□

Note that in the case $f^\circ = \mathbb{1}_{[0,1)}$ neither the test statistic $\widetilde{\mathrm{q}}_k^2$ (defined in (4.4.2)) nor the threshold $\tau_k^{\mathrm{d}}(\alpha)$ (defined in (4.4.5)) depend on the coefficients of the error density φ. Hence, the test $\Delta^{\mathrm{d}}_{k,\alpha/2}$ only depends on the ill-posedness of the model through the dimension parameter k. Aggregating over an appropriate class $\mathcal{K}$ yields a direct max-test $\Delta^{\mathrm{d}}_{\mathcal{K},\alpha/2}$ that is adaptive with respect to both the regularity and the ill-posedness of the model. This observations is formally stated in the next theorem.

Theorem 4.5.5 (Adaptation to ill-posedness). Let $f^\circ = \mathbb{1}_{[0,1)}$. Under the assumptions of Corollary 4.5.3 let $\alpha \in (0,1)$ and consider $\overline{A}_\alpha$ as in (4.4.10). Then for all $A \geq c \cdot C \cdot \overline{A}_\alpha$ and for all $n \in \mathbb{N}$, $n \geq 2$,

$$\sup_{\varphi \in \mathcal{D}} \sup_{a_\bullet \in \mathcal{A}} \mathcal{R}\left(\Delta^{\mathrm{d}}_{\mathcal{K},\alpha/2} \mid \mathcal{E}^{\mathrm{R}}_{a_\bullet}, A\rho^{\mathrm{d}}_{\mathcal{K},a_\bullet}(\delta n)\right) \leq \alpha$$

with $\delta = (1 \vee \log|\mathcal{K}|)^{-1/2}$.

In the next section we provide a lower bound on the minimal adaptive factor δ^{-1} in this situation (i.e. $f = \mathbb{1}_{[0,1)}$).

4.6 Adaptive lower bound

Throughout this section we assume that $f^\circ = \mathbb{1}_{[0,1)}$. The next proposition states general conditions on the class $\mathcal{A}$ under which an adaptive factor δ^{-1} is an unavoidable cost to pay for adaptation over $\mathcal{A}$. The proof of Proposition 4.6.1 is based on a Assouad-type reduction argument and makes use of Lemma D.4.1 in the appendix, which provides a bound on the χ^2-divergence between the null and a mixture over several alternative classes. Inspired by Assouad's cube technique the candidate densities, i.e. the vertices of the hypercubes, are constructed such that, roughly speaking, they are statistically indistinguishable from the null f° while having largest possible $\mathscr{L}^2$-distance.

Proposition 4.6.1 (Adaptive lower bound). Let $\alpha \in (0,1)$ and $\delta \in (0,1]$. Assume there exists a collection of N regularity parameters $\{a^j_\bullet : j \in [\![N]\!]\} \subseteq \mathcal{A}$, where we abbreviate for $j \in [\![N]\!]$

$$\rho^j := \rho_{a^j_\bullet}(\delta n) \qquad \text{with associated optimal dimension parameters} \qquad k^j := k_{a^j_\bullet}(\delta n),$$

such that the following four conditions are satisfied:

(C1) The collection is ordered such that $k^l \leq k^m$ and $\rho^l \leq \delta\rho^m$, whenever $l < m$.

(C2) There exists a finite constant $c_\alpha > 0$ such that $\exp(c_\alpha \delta^{-2}) \leq N\alpha^2$.

(C3) There exists a finite constant $\mathfrak{a}$ such that $2\max_{j \in [\![N]\!]} \|a^j_\bullet\|^2_{\ell^2(\mathbb{N})} \leq \mathfrak{a}$.

(C4) There exists a constant $\eta \in (0,1]$ such that

$$\eta \leq \min_{j \in [\![N]\!]} \frac{(a^j_{k^j})^2 \wedge \frac{\nu^2_{k^j}}{\delta n}}{(a^j_{k^j})^2 \vee \frac{\nu^2_{k^j}}{\delta n}} = \min_{j \in [\![N]\!]} \frac{(a^j_{k^j})^2 \wedge \frac{\nu^2_{k^j}}{\delta n}}{(\rho^j)^2}.$$

Then, with

$$\underline{A}^2_\alpha := \eta\left(\mathrm{R}^2 \wedge \sqrt{\log(1+\alpha^2)} \wedge \mathfrak{a}^{-1} \wedge \sqrt{c_\alpha}\right)$$

we obtain for all $A \in [0, \underline{A}_\alpha]$

$$\inf_{\Delta} \sup_{a_\bullet \in \mathcal{A}} \mathcal{R}\left(\Delta \mid \mathcal{E}^{\mathrm{R}}_{a_\bullet}, A\rho_{a_\bullet}(\delta n)\right) \geq 1 - \alpha.$$

Remark 4.6.2 (Conditions of Proposition 4.6.1). *Let us briefly discuss the conditions of Proposition 4.6.1. Under (C1) the collection of regularity parameters $\mathcal{A}$ is rich enough to make adaptation unavoidable, i.e. it contains distinguishable elements resulting in significantly (measured in terms of δ) different radii. (C2) is a bound for the maximal size of an unavoidable adaptive factor. (C3) guarantees that the candidates constructed in the reduction scheme of the proof are indeed densities. The condition (C4) relates the behaviour of the sequences $\varphi_\bullet$ and $a^j_\bullet$ and essentially guarantees an optimal balance of the bias and the variance term in the dimension k^j uniformly over all $j \in [\![N]\!]$. Moreover, for all regularity and ill-posedness examples considered in Illustration 4.2.5 condition (C4) holds uniformly for all $n \in \mathbb{N}$. We shall emphasise that the optimal dimensions k^j and the corresponding radii ρ^j, for which (C1) and (C4) are stated, are determined in terms of the effective sample size δn.* □

Proof of Proposition 4.6.1. **Reduction Step.** To prove a lower bound for the testing radius we reduce the risk of a test to a distance between probability measures. Denote $\mathbb{P}_0 = \mathbb{P}_{f^\circ}$, let $\mathbb{P}_{1,m}$, specified below, be a mixing measure over the $\underline{A}_\alpha\rho^m$-separated alternative and consider the uniform mixture $\mathbb{P}_1 := \frac{1}{N}\sum_{m\in[\![N]\!]}\mathbb{P}_{1,m}$ over all $m \in [\![N]\!]$. The risk can be lower bounded by applying a classical reduction argument as follows

$$\begin{aligned}
\inf_{\Delta} \sup_{a_\bullet \in \mathcal{A}} \mathcal{R}\left(\Delta \mid \mathcal{E}^{\mathrm{R}}_{a_\bullet}, A\rho_{a_\bullet}(\delta n)\right) &\geq \inf_{\Delta} \max_{m\in[\![N]\!]} \mathcal{R}\left(\Delta \mid \mathcal{E}^{\mathrm{R}}_{a^m_\bullet}, A\rho^m\right) \\
&\geq \inf_{\Delta}\left\{\mathbb{P}_0(\Delta = 1) + \max_{m\in[\![N]\!]} \sup_{f-f^\circ\in\mathcal{E}^{\mathrm{R}}_{a^m_\bullet}\cap\mathcal{L}^2_{\rho^m}} \mathbb{P}_f(\Delta = 0)\right\} \\
&\geq \inf_{\Delta}\left\{\mathbb{P}_0(\Delta = 1) + \frac{1}{N}\sum_{m\in[\![N]\!]} \sup_{f-f^\circ\in\mathcal{E}^{\mathrm{R}}_{a^m_\bullet}\cap\mathcal{L}^2_{\rho^m}} \mathbb{P}_f(\Delta = 0)\right\} \\
&\geq \inf_{\Delta}\left\{\mathbb{P}_0(\Delta = 1) + \frac{1}{N}\sum_{m\in[\![N]\!]} \mathbb{P}_{1,m}(\Delta = 0)\right\} \\
&\geq \inf_{\Delta}\left\{\mathbb{P}_0(\Delta = 1) + \mathbb{P}_1(\Delta = 0)\right\} \\
&= 1 - \mathrm{TV}(\mathbb{P}_0, \mathbb{P}_1) \\
&\geq 1 - \sqrt{\frac{\chi^2(\mathbb{P}_0, \mathbb{P}_1)}{2}},
\end{aligned}$$

where TV denotes the total variation distance and χ^2 the χ^2-divergence.
Definition of the mixture. For each $m \in [\![N]\!]$ we mix the Fourier coefficients uniformly over the vertices of a hypercube contained in the corresponding alternative. For $m \in [\![N]\!]$ we define the coefficients of $\theta^m_\bullet = (\theta^m_j)_{j\in\mathbb{N}} \in \ell^2(\mathbb{N})$ by

$$\theta^m_j := \begin{cases} \frac{\underline{A}_\alpha\rho^m}{\nu^2_{k^m}}\,|\varphi_j|^{-2} & \text{for } j \in [\![k^m]\!] \\ 0 & \text{otherwise} \end{cases}$$

and the candidate functions for the sign vectors $\tau \in \{\pm\}^{k^m}$ by

$$f^{m,\tau} := e_0 + \sum_{|j|\in[\![k^m]\!]} \tau_{|j|}\theta^m_{|j|}e_j.$$

Now, we can define the mixing measures

$$\mathbb{P}_{1,m} := \frac{1}{2^{k^m}} \sum_{\tau \in \{\pm\}^{k^m}} \mathbb{P}_{f^{m,\tau}}, \qquad m \in [\![N]\!].$$

Let us verify that for each $m \in [\![N]\!]$ the collection $\left\{ f^{m,\tau} : \tau \in \{\pm\}^{k^m} \right\}$ is a set of densities contained in the alternative $\mathcal{E}^{\mathrm{R}}_{a^m_\bullet} \cap \mathscr{L}^2_{\rho^m}$, i.e. that $\mathbb{P}_{1,m}$ is indeed supported on the alternative.

1. $\|f^{m,\tau}\|_2^2 = 2\,\|\theta^m_\bullet\|^2_{\ell^2(\mathbb{N})} + 1 < \infty$, satisfied by construction. $(\in \mathscr{L}^2)$

2. $f^{m,\tau}_j = \overline{f^{m,\tau}_{-j}}$, satisfied since θ and τ are vectors in $\mathbb{R}^{\mathbb{N}}$ resp. $\mathbb{R}^{k^m}$, hence (real-valued)

$$f^{m,\tau}_j = \tau_{|j|}\theta^m_{|j|} = \overline{\tau_{|j|}\theta^m_{|j|}} = \overline{f^{m,\tau}_{-j}}.$$

3. $f^{m,\tau}_0 = 1$ satisfied by construction. (normalized to 1)

4. $\sum_{|j|>0} \left| f^{m,\tau}_j \right| \leq 1$, by the Cauchy-Schwarz inequality, (positive)
since

$$\sum_{|j|>0} \left| f^{m,\tau}_j \right| = 2 \sum_{j \in \mathbb{N}} \left| \theta^m_j \right| \frac{a^m_j}{a^m_j} \leq \sqrt{2 \sum_{j\in\mathbb{N}} (a^m_j)^2} \sqrt{\sum_{j\in\mathbb{N}} (a^m_j)^{-2} \left| \theta^m_j \right|^2}$$
$$\leq \sqrt{\mathfrak{a}} \sqrt{\underline{A}^2_\alpha (\rho^m)^2 a^{-2}_{k^m}} \leq \sqrt{\mathfrak{a}}\sqrt{\mathfrak{a}^{-1}} = 1,$$

where the second last inequality follows as in (5.)

5. $f^{m,\tau} - f^\circ \in \mathcal{E}^{\mathrm{R}}_{a^m_\bullet}$, i.e. $2\sum_{j\in\mathbb{N}} (a^m_j)^{-2} \left|\theta^m_j\right|^2 \leq \mathrm{R}^2$, (smoothness)
satisfied since by the monotonicity of $a^m_\bullet$ we have

$$2\sum_{j\in\mathbb{N}} (a^m_j)^{-2} \left|\theta^m_j\right|^2 = \frac{\underline{A}^2_\alpha (\rho^m)^2}{\nu^4_{k^m}} 2 \sum_{j\in[\![k^m]\!]} |\varphi_j|^{-4} (a^m_j)^{-2}$$
$$\leq \underline{A}^2_\alpha (\rho^m)^2 a^{-2}_{k^m} \leq \eta \mathrm{R}^2 (\rho^m)^2 a^{-2}_{k^m} \leq \mathrm{R}^2.$$

6. $f^{m,\tau} - f^\circ \in \mathscr{L}^2_{\underline{A}_\alpha \rho^m}$, i.e. $\|f^{m,\tau} - f^\circ\|_{\mathscr{L}^2} \geq \underline{A}_\alpha \rho^m$ (separation)
satisfied since

$$\|f^{m,\tau} - f^\circ\|^2_{\mathscr{L}^2} = 2 \sum_{j\in[\![k^m]\!]} \left|\theta^m_j\right|^2 = \frac{\underline{A}^2_\alpha (\rho^m)^2}{\nu^4_{k^m}} 2 \sum_{j\in[\![k^m]\!]} |\varphi_j|^{-4} = \underline{A}^2_\alpha (\rho^m)^2.$$

We collect one more property of the constructed densities, which shows that they are similar enough to be statistically indistinguishable.

7. $\frac{1}{N^2} \sum_{l,m\in[\![N]\!]} \exp\big(2n^2 \sum_{j\in[\![k^m \wedge k^l]\!]} \left|\theta^m_j \varphi_j \theta^l_j \varphi_j\right|^2\big) \leq 1 - 2\alpha^2$ (similarity)
Let us first investigate the argument inside of the exp-function.

$$2n^2 \sum_{j\in[\![k^m\wedge k^l]\!]} \left|\theta^m_j \varphi_j \theta^l_j \varphi_j\right|^2 = n^2 \frac{\underline{A}^2_\alpha (\rho^m)^2}{\nu^4_{k^m}} \frac{\underline{A}^2_\alpha (\rho^l)^2}{\nu^4_{k^l}} 2 \sum_{j\in[\![k^m\wedge k^l]\!]} |\varphi_j|^{-4}$$
$$= n^2 \frac{\underline{A}^4_\alpha (\rho^m)^2 (\rho^l)^2}{\nu^4_{k^m \vee k^l}}$$

Let $l < m$, then by (C1), the definition of $\underline{A}_\alpha$ and the condition (C4) on η we obtain

$$\begin{aligned}
2n^2 \sum_{j\in[\![k^m\wedge k^l]\!]} \left|\theta_j^m\varphi_j\theta_j^l\varphi_j\right|^2 &= n^2\frac{\underline{A}_\alpha^4(\rho^m)^2(\rho^l)^2}{\nu_{k^m}^4} \leq n^2\log(1+\alpha^2)\frac{\eta^2(\rho^m)^2(\rho^l)^2}{\nu_{k^m}^4} \\
&= n^2\log(1+\alpha^2)\frac{\eta^2(\rho^m)^4(\rho^l)^2}{(\rho^m)^2\nu_{k^m}^4} \\
&\leq n^2\log(1+\alpha^2)\frac{(\rho^l)^2}{\delta^2n^2(\rho^m)^2} \\
&= \log(1+\alpha^2)\frac{(\rho^l)^2}{\delta^2(\rho^m)^2} \leq \log(1+\alpha^2).
\end{aligned}$$

The case $l = m$ simply yields

$$\begin{aligned}
2n^2 \sum_{j\in[\![k^m\wedge k^l]\!]} \left|\theta_j^m\varphi_j\theta_j^l\varphi_j\right|^2 &= 2n^2 \sum_{j\in[\![k^m]\!]} \left|\theta_j^m\right|^4 |\varphi_j|^4 = 2n^2\frac{\underline{A}_\alpha^4(\rho^m)^4}{\nu_{k^m}^8} \sum_{j\in[\![k^m]\!]} |\varphi_j|^{-8}|\varphi_j|^4 \\
&= n^2\frac{\underline{A}_\alpha^4(\rho^m)^4}{\nu_{k^m}^4} \leq \frac{c_\alpha}{\delta^2}.
\end{aligned}$$

Finally, combining the two bounds, we have

$$\begin{aligned}
&\frac{1}{N^2}\sum_{l,m\in[\![N]\!]} \exp\big(2n^2 \sum_{j\in[\![k^m\wedge k^l]\!]} \left|\theta_j^m\varphi_j\theta_j^l\varphi_j\right|^2\big) \\
&\leq \frac{1}{N^2}\sum_{l=m\in[\![N]\!]} \exp\big(2n^2 \sum_{j\in[\![k^m\wedge k^l]\!]} \left|\theta_j^m\varphi_j\theta_j^l\varphi_j\right|^2\big) \\
&+ \frac{N(N-1)}{N^2}\sum_{l<m\in[\![N]\!]} \exp\big(2n^2 \sum_{j\in[\![k^m\wedge k^l]\!]} \left|\theta_j^m\varphi_j\theta_j^l\varphi_j\right|^2\big) \\
&\leq \frac{1}{N}\exp(c_\alpha\delta^{-2}) + \frac{N(N-1)}{N^2}\exp(\log(1+\alpha^2)) \\
&\leq \frac{N\alpha^2}{N} + 1 + \alpha^2 = 1 + 2\alpha^2, \qquad (4.6.1)
\end{aligned}$$

where the last inequality is due to (C2).

Bound for the χ^2-divergence. We apply Lemma D.4.1 and obtain

$$\chi^2(\mathbb{P}_1, \mathbb{P}_0) \leq \frac{1}{N^2}\sum_{l,m\in[\![N]\!]} \exp\big(2n^2 \sum_{j\in[\![k^m\wedge k^l]\!]} \left|\theta_j^m\varphi_j\theta_j^l\varphi_j\right|^2\big) - 1 \qquad (4.6.2)$$

Hence, property (7.) (similarity) guarantees that the induced distance between the mixing measure and the null is negligible. Combining (4.6.1) with (4.6.2) and the reduction step proves the assertion. □

Adaptive lower bounds in specific situations. We apply Proposition 4.6.1 to two specific classes of alternatives $\left\{\mathcal{E}_{a_\bullet}^{\mathrm{R}} : a_\bullet \in \mathcal{A}\right\}$. We consider a set $\mathcal{A}$ which is **non-trivial** with respect to either a polynomial decay or an exponential decay, that is,

$$\{(j^{-s})_{j\in\mathbb{N}} : s \in [s_\star, s^\star]\} \subseteq \mathcal{A} \qquad \text{or} \qquad \left\{(e^{-j^s})_{j\in\mathbb{N}} : s \in [s_\star, s^\star]\right\} \subseteq \mathcal{A}$$

for $s_\star < s^\star$ and $s_\star, s^\star > 0$.

Theorem 4.6.3 (Minimal adaptive factor – polynomial decay). Let $\mathcal{A}$ be non-trivial with respect to polynomial decay for some $s^\star > s_\star > \frac{1}{2}$. Let $|\varphi_j| \sim j^{-p}$ for some $p > 1/2$. For $\alpha \in (0,1)$ there exists an $n_\circ \in \mathbb{N}$ and $\underline{A}_\alpha \in (0,\infty)$ such that for all $n \geq n_\circ$ and $A \in [0, \underline{A}_\alpha]$

$$\inf_{\Delta} \sup_{a_\bullet \in \mathcal{A}} \mathcal{R}\left(\Delta \mid \mathcal{E}^{\mathrm{R}}_{a_\bullet}, A\rho_{a_\bullet}(\delta n)\right) \geq 1 - \alpha$$

with $\delta = (1 \vee \log\log n)^{-\frac{1}{2}}$, i.e. δ^{-1} is a lower bound for the minimal adaptive factor over $\mathcal{A}$.

Proof of Theorem 4.6.3. We intend to apply Proposition 4.6.1. To do so, we construct a collection of regularity parameters $\mathcal{A}_N := \{a^m_\bullet \in \mathcal{A} : m \in [\![N]\!]\} \subseteq \mathcal{A}$ such that (C1) – (C4) are satisfied.

Definition of the collection. Recall from Illustration 4.2.5 that the minimax radius in our setting is of order $\rho^2_{a_\bullet}(\delta n) \sim (\delta n)^{-e(s)}$ with the exponent $e(s) := \frac{4s}{4s+4p+1}$. Since $\mathcal{A}$ is non-trivial with respect to polynomial decay, it contains a subset of the form $\{(j^{-s})_{j\in\mathbb{N}} : s \in [s_\star, s^\star]\}$. Due to $e(s) = \frac{4s}{4s+4p+1} = 1 - \frac{4p+1}{4s+4p+1}$ the exponent is monotonically increasing in s, hence the corresponding regularity parameters result in radii with exponents in the interval $[e(s_\star), e(s^\star)] =: [e_\star, e^\star]$. We define a grid of size N on $[e_\star, e^\star]$, which then induces a grid on $[s_\star, s^\star]$. For the step size $d := \frac{e^\star - e_\star}{N}$ let

$$\mathcal{G}_e := \{e^\star - md : m \in \{0,\dots,N-1\}\} \subseteq [e_\star, e^\star]$$

be a grid on $[e_\star, e^\star]$. For $m \in \{0,\dots,N-1\}$ let s_m be defined by the equation

$$e(s_m) = \frac{4s_m}{4s_m + 4p + 1} = e^\star - md,$$

which results in a grid on $[s_\star, s^\star]$,

$$\mathcal{G}_s := \{s_m : e(s_m) = e^\star - md, m \in \{0,\dots,N-1\}\}.$$

Finally, we define our collection of regularity sequences as

$$\mathcal{G}_{a_\bullet} := \{(j^{-s})_{j\in\mathbb{N}} : s \in \mathcal{G}_s\}.$$

Verification of the conditions (C1) – (C4)

(C1) Let n be large enough such that the effective sample size δn is larger than 1. The grid is defined such that

$$\begin{aligned} m > l &\iff e(s_m) = e^\star - md < e^\star - ld = e(s_l) \\ &\iff s_m < s_l \\ &\iff \tfrac{2}{4p+4s_m+1} > \tfrac{2}{4p+4s_l+1} \\ &\iff (\delta n)^{\frac{2}{4p+4s_m+1}} > (\delta n)^{\frac{2}{4p+4s_l+1}}, \end{aligned}$$

which shows that, since by Illustration 4.2.5, $k^m \sim (\delta n)^{\frac{2}{4p+4s_m+1}}$,

$$\lim_{n\to\infty} \frac{k^l}{k^m} < 1.$$

In other words, there exists a $n_{\circ,1} \in \mathbb{N}$ such that for all $n \geq n_{\circ,1}$

$$k^l < k^m, \qquad \text{for } l < m,$$

which verifies the first part of (C1).

To check the second part, we define $N := \lfloor \frac{e^\star - e_\star}{4} \frac{\log(\delta n)}{|\log(\delta)|} \rfloor$, where we assume that n is large enough such that $N \geq 1$. Since for $l < m$,

$$\frac{(\delta n)^{-e(s_l)}}{(\delta n)^{-e(s_m)}} = (\delta n)^{e(s_m)-e(s_l)} = (\delta n)^{(l-m)d},$$

we obtain

$$\delta^{-2} \frac{(\delta n)^{-e(s_l)}}{(\delta n)^{-e(s_m)}} = \exp\left((l-m)d\log(\delta n) - 2\log\delta\right)$$

with

$$\begin{aligned}(m-l)d\log(\delta n) + 2\log\delta &\geq d\log(\delta n) + 2\log(\delta) \\ &\geq 4\frac{|\log(\delta)|}{\log(\delta n)}\log(\delta n) + 2\log(\delta) = 4|\log(\delta)| + 2\log(\delta) > 0.\end{aligned}$$

Hence, due to $(\rho^l)^2 \sim (\delta n)^{e(s_l)}$, we obtain

$$\lim_{n\to\infty} \delta^{-2}\frac{(\rho^l)^2}{(\rho^m)^2} < 1.$$

In other words, there exists a $n_{\circ,2} \in \mathbb{N}$ such that for all $n \geq n_{\circ,2}$

$$\rho^l < \delta\rho^m \qquad \text{for } l < m.$$

(C2) The condition (C2) can be rewritten as

$$c_\alpha \delta^{-2} - \log(N) \leq 2\log(\alpha).$$

It is easily seen that $\delta^2\log(N) \longrightarrow 1$ for $n \to \infty$. Hence, $\log(N) - \frac{1}{2}\delta^{-2} \longrightarrow \infty$ and, thus, there exists a $n_{\circ,3}$ (possibly depending on α) such that

$$c_\alpha \delta^{-2} - \log(N) \leq 2\log(\alpha).$$

with $c_\alpha := \frac{1}{2}$ and, therefore, $(C2)$ is satisfied.

(C3) We observe that

$$\sup_{m\in[\![N]\!]}\sum_{j\in\mathbb{N}}(a_j^m)^2 \leq \sup_{s\in[s_\star,s^\star]}\sum_{j\in\mathbb{N}} j^{-2s} \leq \sum_{j\in\mathbb{N}} j^{-2s_\star} \leq \int_1^\infty x^{-2s_\star}\mathrm{d}x \leq \frac{1}{2s_\star - 1} =: \mathfrak{a},$$

which shows (C3).

(C4) The existence of a constant η satisfying (C4) uniformly over n follows, because for $a_\bullet \sim (j^{-s})_{j\in\mathbb{N}}$ with $s \in [s_\star, s^\star]$ the terms $a_{k_{a_\bullet}}^2$ and $\frac{\nu_{k_{a_\bullet}}^2}{\delta n}$ are of the same order.

□

Theorem 4.6.4 (Minimal adaptive factor – exponential decay). Let $\mathcal{A}$ be non-trivial with respect to exponential decay for some $s^\star > s_\star > 0$. Let $|\varphi_j| \sim j^{-p}$ for some $p > 1/2$. For $\alpha \in (0,1)$ there exists an $n_\circ \in \mathbb{N}$ and $\underline{A}_\alpha \in (0,\infty)$ such that for all $n \geq n_\circ$ and $A \in [0, \underline{A}_\alpha]$

$$\inf_\Delta \sup_{a_\bullet\in\mathcal{A}} \mathcal{R}\left(\Delta \mid \mathcal{E}_{a_\bullet}^{\mathrm{R}}, A\rho_{a_\bullet}(\delta n)\right) \geq 1 - \alpha$$

with $\delta = (1 \vee \log\log\log n)^{-1/2}$, i.e. δ^{-1} is a lower bound for the minimal adaptive factor over $\mathcal{A}$.

Proof of Theorem 4.6.4. We intend to apply Proposition 4.6.1. To do so, we construct a collection of regularity parameters $\mathcal{A}_N := \{a_\bullet^m \in \mathcal{A} : m \in [\![N]\!]\} \subseteq \mathcal{A}$ such that (C1)–(C4) are satisfied.
Definition of the collection. Recall from Illustration 4.2.5 that the minimax radius in our setting is of order $\rho^2_{a_\bullet}(\delta n) \sim \frac{(\log \delta n)^{e(s)}}{\delta n}$ with the exponent $e(s) := \frac{2p+1/2}{s}$. Since $\mathcal{A}$ is non-trivial with respect to exponential decay, it contains a subset of the form $\{(e^{-j^s})_{j\in\mathbb{N}} : s \in [s_\star, s^\star]\}$. The exponent is monotonically decreasing in s, hence the corresponding regularity parameters result in radii with exponents in the interval $[e(s^\star), e(s_\star)] =: [e_\star, e^\star]$. We define a grid of size N on $[e_\star, e^\star]$, which then induces a grid on $[s_\star, s^\star]$. For $d := \frac{e^\star - e_\star}{N}$ let

$$\mathcal{G}_e := \{e_\star + md : m \in \{0, \dots, N-1\}\} \subseteq [e_\star, e^\star]$$

be a grid on $[e_\star, e^\star]$. For $m \in \{0, \dots, N-1\}$ let s_m be defined by the equation

$$e(s_m) = \frac{2p + 1/2}{s_m} = e^\star + md,$$

which results in a grid on $[s_\star, s^\star]$

$$\mathcal{G}_s := \{s_m : e(s_m) = e_\star + md, m \in \{0, \dots, N-1\}\}.$$

Finally, we define our collection of regularity sequences as

$$\mathcal{G}_{a_\bullet} := \left\{(e^{-j^s})_{j\in\mathbb{N}} : s \in \mathcal{G}_s\right\}.$$

Verification of the conditions (C1)–(C4)

(C1) Let n be large enough such that the effective sample size δn is larger than 1. The grid is defined such that

$$\begin{aligned} m > l &\iff e(s_m) = e^\star + md > e^\star + ld = e(s_l) \\ &\iff \tfrac{1}{s_m} > \tfrac{1}{s_l} \\ &\iff (\delta n)^{\frac{1}{s_m}} > (\delta n)^{\frac{1}{s_l}}, \end{aligned}$$

which shows that, since by Illustration 4.2.5, $k^m \sim (\delta n)^{\frac{1}{s_m}}$,

$$\lim_{n\to\infty} \frac{k^l}{k^m} < 1.$$

In other words, there exists a $n_{\circ,1} \in \mathbb{N}$ such that for all $n \geq n_{\circ,1}$

$$k^l < k^m \qquad \text{for } l < m,$$

which verifies the first part of (C1).
To check the second part, we define $N := \lfloor \frac{e^\star - e_\star}{4} \frac{\log\log(\delta n)}{|\log(\delta)|} \rfloor$, where we assume that n is large enough such that $N \geq 1$. Since for $l < m$,

$$\frac{\frac{1}{\delta n}(\log(\delta n))^{e(s_l)}}{\frac{1}{\delta n}(\log(\delta n))^{e(s_m)}} = (\log(\delta n))^{e(s_l)-e(s_m)} = (\log(\delta n))^{(l-m)d}$$

we obtain

$$\delta^{-2} \frac{\frac{1}{\delta n}(\log(\delta n))^{e(s_l)}}{\frac{1}{\delta n}(\log(\delta n))^{e(s_m)}} = \exp\left((l-m)d\log\log(\delta n) - 2\log\delta\right)$$

with

$$
\begin{aligned}
(m-l)d\log\log(\delta n)+2\log\delta &\geq d\log\log(\delta n)+2\log(\delta)\\
&\geq 4\frac{|\log(\delta)|}{\log\log(\delta n)}\log\log(\delta n)+2\log(\delta)\\
&= 4\,|\log(\delta)|+2\log(\delta)>0.
\end{aligned}
$$

Hence, due to $(\rho^l)^2 \sim \frac{(\log \delta n)^{e(s_l)}}{\delta n}$, we obtain

$$\lim_{n\to\infty}\delta^{-2}\frac{(\rho^l)^2}{(\rho^m)^2}<1.$$

In other words, there exists a $n_{\circ,2}\in\mathbb{N}$ such that for all $n\geq n_{\circ,2}$

$$\rho^l<\delta\rho^m \qquad \text{for } l<m.$$

(C2) The condition (C2) can be rewritten as

$$c_\alpha\delta^{-2}-\log(N)\leq 2\log(\alpha).$$

It is easily seen that $\delta^2\log(N)\longrightarrow 1$ for $n\to\infty$. Hence, $\log(N)-\frac{1}{2}\delta^{-2}\longrightarrow\infty$ and, thus, there exists a $n_{\circ,3}$ (possibly depending on α) such that

$$c_\alpha\delta^{-2}-\log(N)\leq 2\log(\alpha).$$

with $c_\alpha=\frac{1}{2}$ and, therefore, $(C2)$ is satisfied.

(C3) We observe that

$$\sup_{m\in[\![N]\!]}\sum_{j\in\mathbb{N}}(a_j^m)^2\leq\sup_{s\in[s_\star,s^\star]}\sum_{j\in\mathbb{N}}e^{-2j^s}\leq\sum_{j\in\mathbb{N}}e^{-2j^{s_\star}}\leq\int_0^\infty e^{-2x^{s_\star}}\mathrm{d}x,$$

where we introduce the change of variables $y=2x^{s_\star}, \mathrm{d}x=1/s_\star(1/2)^{1/s_\star}y^{1/s_\star}\mathrm{d}y$

$$
\begin{aligned}
&=\left(\frac{1}{2}\right)^{\frac{1}{s_\star}}\frac{1}{s_\star}\int_0^\infty y^{\frac{1}{s}}e^{-y}\mathrm{d}y\\
&=\left(\frac{1}{2}\right)^{\frac{1}{s_\star}}\frac{1}{s_\star}\Gamma\left(\frac{1}{s_\star}\right)=\left(\frac{1}{2}\right)^{\frac{1}{s_\star}}\Gamma\left(\frac{1}{s_\star}+1\right)=:\mathfrak{a},
\end{aligned}
$$

which shows (C3).

(C4) The existence of a constant η satisfying (C4) uniformly over n follows, because for $a_\bullet\sim(e^{-j^s})_{j\in\mathbb{N}}$ with $s\in[s_\star,s^\star]$ the terms $a^2_{k_{a_\bullet}}$ and $\frac{\nu^2_{k_{a_\bullet}}}{\delta n}$ are of the same order.

□

Comparing Theorem 4.6.3 and Theorem 4.6.4 with Illustration 4.3.6 (for the indirect test) and Illustration 4.5.4 (for the direct test) shows that the adaptive factors that we obtain are minimal. Indeed, in the ordinary smooth – mildly ill-posed model both the direct and the indirect max-test face a deterioration by a $\sqrt{\log\log n}$-factor, which Theorem 4.6.3 shows to be unavoidable. In the more restrictive setting of super smoothness and mild ill-posedness both tests feature a $\sqrt{\log\log\log n}$-factor, which is unavoidable due to Theorem 4.6.4. In the ordinary smooth – severely ill-posed model there is no loss for adaptation visible in the testing radius. Finally, let us comment on the fact that Theorem 4.6.3 and Theorem 4.6.4 at first glance only provide asymptotic results since they require n to be sufficiently large. This is quite a natural assumption since the adaptive factor δ^{-1} only has an effect if $\sqrt{\log\log n}>1$ (for polynomial decay) or if $\sqrt{\log\log\log n}>1$ (for exponential decay), which clearly only occurs for n large enough.

Appendix D

Auxiliary results

D.1 Preliminaries

The next two assertions, a concentration inequality for canonical U-statistics and a Bernstein inequality, provide our key arguments in order to control the deviation of the test statistics. The first assertion is a reformulation of Theorem 3.4.8 in Gine and Nickl [2015].

Proposition D.1.1 (Concentration inequality for U-statistics). Let $n \geq 2$ and let $\{Y_l\}_{l=1}^n$ be independent and identically distributed $[0,1)$-valued random variables. Let $h : [0,1)^2 \longrightarrow \mathbb{R}$ be a bounded symmetric kernel, i.e. $h(y, \tilde{y}) = h(\tilde{y}, y)$ for all $y, \tilde{y} \in [0,1)$, fulfilling in addition

$$\mathbb{E}(h(Y_1, y_2)) = 0 \qquad \forall y_2 \in [0,1). \tag{D.1.1}$$

Let A, B, C and D be real numbers such that

$$\begin{aligned} \sup_{y_1, y_2 \in [0,1)} |h(y_1, y_2)| &\leq A, \\ \sup_{y_2 \in [0,1)} \mathbb{E}h^2(Y_1, y_2) &\leq B^2, \\ \mathbb{E}h^2(Y_1, Y_2) &\leq C^2, \\ \sup \left\{ \mathbb{E}(h(Y_1, Y_2)\zeta(Y_1)\xi(Y_2)), \mathbb{E}\zeta^2(Y_1) \leq 1, \mathbb{E}\xi^2(Y_2) \leq 1 \right\} &\leq D. \end{aligned} \tag{D.1.2}$$

Then, the real-valued canonical U-statistic

$$\mathrm{U}_n = \frac{1}{n(n-1)} \sum_{\substack{l,m \in [\![n]\!] \\ l \neq m}} h(Y_l, Y_m)$$

satisfies for all $x \geq 0$

$$\mathbb{P}\left(\mathrm{U}_n \geq 8\frac{C}{n} x^{1/2} + 13\frac{D}{n} x + 261\frac{B}{n^{3/2}} x^{3/2} + 343\frac{A}{n^2} x^2 \right) \leq \exp(1 - x).$$

The following version of Bernstein's inequality can directly be deduced from Theorem 3.1.7. in Gine and Nickl [2015].

Proposition D.1.2 (Bernstein's inequality).
Let $\{Z_j\}_{j=1}^n$ be independent random variables with $|Z_j| \leq \mathrm{b}$ almost surely and $\mathbb{E}\,|Z_j|^2 \leq \mathrm{v}$

for all $j \in [\![n]\!]$. Then for all $x > 0$ and $n \geq 1$, we have

$$\mathbb{P}\left(\frac{1}{n}\sum_{j\in[\![n]\!]}(Z_j - \mathbb{E}Z_j) \geq \sqrt{\frac{2\mathrm{v}x}{n}} + \frac{\mathrm{b}x}{3n}\right) \leq \exp(-x).$$

Preliminaries. We assume throughout this section that Y and $\{Y_j\}_{j=1}^n$ are independent and identically distributed with density $g = f \circledast \varphi \in \mathscr{L}^2$ with respect to the Lebesgue measure.

Rewriting the condition on D. Recall that by $\mathscr{L}^2(g)$ we denote the set of Borel-measurable functions $\xi : [0,1) \to \mathbb{R}$ with $\|\xi\|^2_{\mathscr{L}^2(g)} := \int_0^1 \xi^2(x)g(x)\mathrm{d}x < \infty$. The associated inner product is given by $\langle \xi, \zeta\rangle_{\mathscr{L}^2(g)} := \int_0^1 \xi(x)\zeta(x)g(x)\mathrm{d}x$ for $\xi, \zeta \in \mathscr{L}^2(g)$. We express the condition on D as the norm of an operator from $\mathscr{L}^2(g)$ into itself. Let $h : [0,1)^2 \longrightarrow \mathbb{R}$ be a bounded kernel, i.e. $\|h\|_{\mathscr{L}^\infty} := \sup_{y_1,y_2\in[0,1)} |h(y_1,y_2)| < \infty$. Consider the integral operator

$$\begin{aligned} H : \mathscr{L}^2(g) &\longrightarrow \mathscr{L}^2(g) \\ \xi &\longmapsto H\xi \end{aligned} \tag{D.1.3}$$

with $H\xi(s) := \int_0^1 h(t,s)\xi(t)g(t)\mathrm{d}t = \mathbb{E}h(Y,s)\xi(Y)$ for $s \in [0,1)$. The operator has the following properties.

1. H is **well-defined**, i.e. $H\xi \in \mathscr{L}^2(g)$, since

$$\begin{aligned} \|H\xi\|^2_{\mathscr{L}^2} &= \int_0^1 |H\xi(x)|^2 g(x)\mathrm{d}x = \int_0^1 \left(\int_0^1 h(t,x)\xi(t)g(t)\mathrm{d}t\right)^2 g(x)\mathrm{d}x \\ &\leq \int_0^1 \left(\int_0^1 h^2(t,x)g(t)\mathrm{d}t\right)\left(\int_0^1 \xi^2(t)g(t)\mathrm{d}t\right) g(x)\mathrm{d}x \\ &= \|\xi\|_{\mathscr{L}^2(g)} \mathbb{E}\,|h(Y_1,Y_2)|^2 , \end{aligned} \tag{D.1.4}$$

   where we applied the Cauchy-Schwarz inequality in the second line. Hence, $\|H\xi\|^2_{\mathscr{L}^2}$ is finite, since h is bounded and $\xi \in \mathscr{L}^2(g)$.

2. H is **linear**, i.e. for $\xi, \zeta \in \mathscr{L}^2(g)$, $\lambda \in \mathbb{R}$, $H(\lambda\xi + \zeta) = \lambda H\xi + H\zeta$, since integrals are linear.

3. The **operator norm** of H is bounded by $\|h\|_{\mathscr{L}^\infty}$, since due to (D.1.4) we have

$$\|H\|_{\mathscr{L}^2(g)\to\mathscr{L}^2(g)} := \sup\left\{\|H\xi\|_{\mathscr{L}^2(g)} : \|\xi\|_{\mathscr{L}^2} \leq 1\right\} \leq \|h\|_{\mathscr{L}^\infty} .$$

4. The **operator norm** can be written as

$$\|H\|_{\mathscr{L}^2(g)\to\mathscr{L}^2(g)} = \sup\left\{\mathbb{E}(h(Y_1,Y_2)\zeta(Y_1)\xi(Y_2), \mathbb{E}\zeta^2(Y_1) \leq 1, \mathbb{E}\xi^2(Y_2) \leq 1\right\}. \tag{D.1.5}$$

   Indeed, note that the operator norm satisfies

$$\begin{aligned} \|H\|_{\mathscr{L}^2(g)\to\mathscr{L}^2(g)} &= \sup\left\{\|H\zeta\|_{\mathscr{L}^2(g)} : \|\zeta\|_{\mathscr{L}^2} \leq 1\right\} \leq \|h\|_{\mathscr{L}^\infty} \\ &= \sup\left\{\sqrt{\mathbb{E}(H\zeta)^2(Y_2)} : \|\zeta\|_{\mathscr{L}^2} \leq 1\right\} \\ &= \sup\left\{\sqrt{\mathbb{E}h^2(Y_1,Y_2)\zeta^2(Y_1)} : \|\zeta\|_{\mathscr{L}^2} \leq 1\right\} \end{aligned}$$

   The claim then follows by applying the Cauchy-Schwarz inequality to

$$\mathbb{E}(h(Y_1,Y_2)\zeta(Y_1)\xi(Y_2)) \leq \left(\mathbb{E}h^2(Y_1,Y_2)\zeta^2(Y_1)\right)^{1/2}\left(\mathbb{E}\xi^2(Y_2)\right)^{1/2}$$

   and taking the supremum over $\mathbb{E}\zeta^2(Y_1) \leq 1$ and $\mathbb{E}\xi^2(Y_2) \leq 1$.

5. Let the kernel h by symmetric, then H is **self-adjoint**. Let $\zeta, \xi \in \mathscr{L}^2(g)$, then we have

$$\begin{aligned}\langle \zeta, H\xi\rangle_{\mathscr{L}^2(g)} &= \int_0^1 \zeta(x)(H\xi)(x)\mathrm{d}x = \int_0^1\int_0^1 h(z,x)\xi(z)g(z)\mathrm{d}zg(x)\zeta(x)\mathrm{d}x \\ &= \int_0^1\int_0^1 \zeta(x)h(x,z)g(x)\mathrm{d}xg(z)\xi(z)\mathrm{d}z = \int_0^1 \xi(z)(H\zeta)(z)g(z)\mathrm{d}z = \langle H\zeta, \xi\rangle_{\mathscr{L}^2(g)},\end{aligned}$$

where we used Fubini's Theorem and the symmetry of h.

Hence, for h bounded and symmetric, H is a linear, bounded and self-adjoint operator. Thus, we can write the operator norm of H as (see Theorem V.5.7 of Werner [2006])

$$\|H\|_{\mathscr{L}^2(g)\to\mathscr{L}^2(g)} = \sup\left\{\left|\langle H\xi, \xi\rangle_{\mathscr{L}^2(g)}\right| : \|\xi\|_{\mathscr{L}^2(g)} \le 1\right\}. \tag{D.1.6}$$

Note that due to (D.1.4), we can always use $D := C$. Under an additional assumption we are, however, able to achieve a sharper bound. For this we recall some properties of the discrete convolution in the next paragraph.

Discrete convolution. Recall that for $p \ge 1$ we denote by $\ell^p := \ell^p(\mathbb{Z})$ the Banach space of complex-valued sequences over $\mathbb{Z}$ endowed with its usual ℓ^p-norm given by $\|a_\bullet\|_{\ell^p} := \left(\sum_{j\in\mathbb{Z}}|a_j|^p\right)^{1/p}$ for $a_\bullet := (a_j)_{j\in\mathbb{Z}} \in \mathbb{C}^{\mathbb{Z}}$. In the case $p = 2$, the space ℓ^2 is a Hilbert space and the ℓ^2-norm is induced by its usual inner product $\langle a_\bullet, b_\bullet\rangle_{\ell^2} := \sum_{j\in\mathbb{Z}} a_j\overline{b_j}$ for all $a_\bullet, b_\bullet \in \ell^2$. For each sequence $a_\bullet \in \ell^1$, we define the discrete convolution operator

$$\begin{aligned} a_\bullet * : \ell^2 &\longrightarrow \ell^2 \\ b_\bullet &\longmapsto (a_\bullet * b_\bullet)\end{aligned}$$

with $(a_\bullet * b_\bullet)_j := \sum_{l\in\mathbb{Z}} a_{j-l}b_l$. The following propositions collect some properties of the discrete convolution operator. Similar results hold for the (continuous) convolution operator, for which the proofs can be found in Werner [2006] (Example on p.348). Our proofs for the discrete case are similar to those in Werner [2006], but we state them here for completeness.

Proposition D.1.3 (Properties of the discrete convolution operator).
Let $a_\bullet \in \ell^2 \cap \ell^1$.

1. The operator $a_\bullet*$ is well defined, i.e. $a_\bullet * b_\bullet \in \ell^2$ for all $b_\bullet \in \ell^2$.
2. The operator $a_\bullet*$ is linear and continuous.
3. The operator norm satisfies $\|a_\bullet*\|_{\ell^2\to\ell^2} \le \|a_\bullet\|_{\ell^1}$.
4. If $a_j = \overline{a_{-j}}$ for all $j \in \mathbb{Z}$, then $a_\bullet*$ is self-adjoint.
5. If $a_j = \overline{a_{-j}}$ for all $j \in \mathbb{Z}$, then $|\langle a_\bullet * b_\bullet, b_\bullet\rangle_{\ell^2}| \le \|a_\bullet*\|_{\ell^2\to\ell^2}\|b_\bullet\|_{\ell^2}^2$ for all $b \in \ell^2$.
6. For $b \in \ell^2$ we have $|\langle a_\bullet * b_\bullet, b_\bullet\rangle_{\ell^2}| \le \|a_\bullet\|_{\ell^1}\|b_\bullet\|_{\ell^2}^2$.

Proof of Proposition D.1.3. 1. We first show that each coefficient is finite.

$$\left|\sum_{l\in\mathbb{Z}} a_{j-l}b_l\right|^2 \le \left|\sum_{l\in\mathbb{Z}}|a_{j-l}|\,|b_l|\right|^2 \le \left(\sum_{l\in\mathbb{Z}}|a_{j-l}|^2\right)\left(\sum_{l\in\mathbb{Z}}|b_l|^2\right) = \|a_\bullet\|_{\ell^2}^2\|b_\bullet\|_{\ell^2}^2 < \infty.$$

Next, we prove that $a_\bullet * b_\bullet \in \ell^2$ for all $b_\bullet \in \ell^2$. We first observe that for $j \in \mathbb{Z}$ by introducing $m = j - l$ we obtain

$$|(a_\bullet * b_\bullet)_j| = \left|\sum_{l\in\mathbb{Z}} a_{j-l} b_l\right|^2 \leq \left|\sum_{m\in\mathbb{Z}} |a_m|\,|b_{j-m}|\right|^2 = \left(\sum_{m\in\mathbb{Z}} \sqrt{|a_m|}\sqrt{|a_m|}\,|b_{j-m}|\right)^2$$
$$\leq \left(\sum_m |a_m|\right)\left(\sum_m |a_m|\,|b_{j-m}|^2\right) = \|a_\bullet\|_{\ell^1} \sum_{m\in\mathbb{Z}} |a_m|\,|b_{j-m}|^2$$

and, hence,

$$\begin{aligned}\sum_{j\in\mathbb{Z}} |(a_\bullet * b_\bullet)_j|^2 &\leq \|a_\bullet\|_{\ell^1} \sum_{j\in\mathbb{Z}}\sum_{m\in\mathbb{Z}} |a_m|\,|b_{j-m}|^2 \\ &= \|a_\bullet\|_{\ell^1} \sum_{m\in\mathbb{Z}} \left(\sum_{j\in\mathbb{Z}} |b_{j-m}|^2\right) |a_m| \\ &= \|a_\bullet\|_{\ell^1} \|b_\bullet\|_{\ell^2}^2 \sum_{m\in\mathbb{Z}} |a_m| \\ &= \|a_\bullet\|_{\ell^1}^2 \|b_\bullet\|_{\ell^2}^2 < \infty.\end{aligned}$$

2. To check linearity let $b_\bullet, c_\bullet \in \ell^2(\mathbb{Z})$, $\lambda, \mu \in \mathbb{R}$. Then,

$$\begin{aligned}(a_\bullet * (\lambda b_\bullet + \mu c_\bullet))_j &= \sum_{l\in\mathbb{Z}} a_{j-l}(\lambda b_l + \mu c_l) \\ &= \lambda \sum_{l\in\mathbb{Z}} a_{j-l} b_l + \mu \sum_{l\in\mathbb{Z}} a_{j-l} c_l = \lambda(a_\bullet * b_\bullet)_j + \mu(a_\bullet * c_\bullet)_j.\end{aligned}$$

A linear operator between normed spaces $T : \mathcal{X} \to \mathcal{Y}$ is continuous if and only if there exists an $M > 0$ such that $\|Tx\|_{\mathcal{Y}} \leq M\,\|x\|_{\mathcal{X}}$ for all $x \in \mathcal{X}$ (see e.g. Werner [2006], Theorem II.1.2). In our case $T = a_\bullet *$, $\mathcal{X} = \mathcal{Y} = \ell^2(\mathbb{Z})$, we have due to (1.)

$$\|a_\bullet * b_\bullet\|_{\ell^2} \leq \|a_\bullet\|_{\ell^1}\,\|b_\bullet\|_{\ell^2}\,,$$

hence $a_\bullet *$ is linear and continuous.

3. From (1.) it follows that

$$\|a_\bullet *\|_{\ell^2\to\ell^2} = \sup_{\|b_\bullet\|_{\ell^2}=1} \|a_\bullet * b_\bullet\|_{\ell^2} = \sup_{b_\bullet\neq 0_\bullet} \frac{\|a_\bullet * b_\bullet\|_{\ell^2}}{\|b_\bullet\|_{\ell^2}} \leq \|a_\bullet\|_{\ell^1}$$

4. Let $b_\bullet, c_\bullet \in \ell^2(\mathbb{Z})$. We have

$$\begin{aligned}\langle a_\bullet * b_\bullet, c_\bullet\rangle_{\ell^2} &= \sum_{j\in\mathbb{Z}}\sum_{l\in\mathbb{Z}} a_{j-l} b_l \overline{c_j} \\ &= \sum_{l\in\mathbb{Z}}\sum_{j\in\mathbb{Z}} \overline{a_{l-j} c_j}\, b_l = \sum_{l\in\mathbb{Z}} \overline{(a_\bullet * c_\bullet)_l}\, b_l = \langle b_\bullet, a_\bullet * c_\bullet\rangle_{\ell^2},\end{aligned}$$

which implies the self-adjointness of $a_\bullet *$.

5. A linear, continuous and self-adjoint operator $T : \mathcal{H} \longrightarrow \mathcal{H}$ between Hilbert spaces satisfies $\|T\|_{\mathcal{H}\to\mathcal{H}} = \sup_{\|x\|_{\mathcal{H}}\leq 1} |\langle Tx, x\rangle_{\mathcal{H}}| = \sup_{x\neq 0_{\mathcal{H}}} \frac{|\langle Tx,x\rangle_{\mathcal{H}}|}{\|x\|_{\mathcal{H}}^2}$. (cp. Werner [2006], V.5.7). Since $a_\bullet$ is linear, continuous and self-adjoint by (2.) and (4.), we obtain for all $b_\bullet \in \ell^2 \setminus \{0\}$ that

$$\|a_\bullet *\|_{\ell^2\to\ell^2} \geq \frac{|\langle a_\bullet * b_\bullet, b_\bullet\rangle_{\ell^2}|}{\|b_\bullet\|_{\ell^2}^2}.$$

Hence,

$$|\langle a_\bullet * b_\bullet, b_\bullet\rangle_{\ell^2}| \leq \|a_\bullet *\|_{\ell^2\to\ell^2}\,\|b_\bullet\|_{\ell^2}^2$$

6. If $a_j = \overline{a_{-j}}$ for all $j \in \mathbb{Z}$, then the assertion immediately follows from (5.) combined with (3.).
   For arbitrary $a_\bullet \in \ell^1(\mathbb{Z}) \cap \ell^2(\mathbb{Z})$, note that using similar calculations as in (1.) we obtain

$$
\begin{aligned}
|\langle a_\bullet * b_\bullet, b_\bullet \rangle_{\ell^2}|^2 &= \left| \sum_{j\in\mathbb{Z}} \sum_{l\in\mathbb{Z}} a_{j-l} b_l \overline{b_j} \right|^2 \\
&\leq \left(\sum_{j\in\mathbb{Z}} \left| \overline{b_j} \right|^2 \right) \left(\sum_{j\in\mathbb{Z}} \left| \sum_{l\in\mathbb{Z}} a_{j-l} b_l \right|^2 \right) \\
&\leq \|b_\bullet\|_{\ell^2}^2 \, \|a_\bullet\|_{\ell^1} \sum_{j\in\mathbb{Z}} \sum_{m\in\mathbb{Z}} |a_m| \, |b_{j-m}|^2 \\
&\leq \|b_\bullet\|_{\ell^2}^2 \, \|a_\bullet\|_{\ell^1} \sum_{m\in\mathbb{Z}} |a_m| \sum_{j\in\mathbb{Z}} |b_{j-m}|^2 \\
&= \|b_\bullet\|_{\ell^2}^4 \, \|a_\bullet\|_{\ell^1}^2 \, ,
\end{aligned}
$$

   which completes the proof.

□

Recall that the real density $g \in \mathscr{L}^2$ of the observations satisfies $g = f \circledast \varphi$ with both f and φ belonging to $\mathscr{L}^2$. Consequently, the Fourier coefficients $g_\bullet = (g_j)_{j\in\mathbb{Z}}$ belong to both ℓ^2, since $\|g\|_{\mathscr{L}^2} = \|g_\bullet\|_{\ell^2}$ by Parseval's identity, and to ℓ^1 due to the convolution theorem. Indeed, since $g_j = f_j \varphi_j$ for all $j \in \mathbb{Z}$, we obtain $\|g_\bullet\|_{\ell^1} \leq \|f_\bullet\|_{\ell^2} \|\varphi_\bullet\|_{\ell^2} < \infty$ due to the Cauchy-Schwarz inequality.

Corollary D.1.4 (Discrete convolution with the coefficients of a density). Let $g \in \mathcal{D}$ with Fourier coefficients $g_\bullet = (g_j)_{j\in\mathbb{Z}}$ in $\ell^1 \cap \ell^2$. Then, the discrete convolution operator $g_\bullet * : \ell^2 \to \ell^2$ is linear, bounded and self-adjoint.

Proof of Corollary D.1.4. The assumption $g \in \mathscr{L}^2$ implies $g_\bullet \in \ell^2$ due to Parseval's identity. The density g is real-valued, hence, we have $g_j = \overline{g_{-j}}$ for all $j \in \mathbb{Z}$ (cp. (4.2.3)). The claim then immediately follows from Proposition D.1.3. □

Under an additional assumption on the space (of real-valued) functions $\mathscr{L}^2(g)$, the operator $g_\bullet *$ is a non-negative. Hence, there exists an operator $(g_\bullet *)^{1/2}$ such that $\left\| (g_\bullet *)^{1/2} \xi_\bullet \right\|_{\ell^2}^2 = \langle g_\bullet * \xi_\bullet, \xi_\bullet \rangle_{\ell^2}$. This is used frequently in the proofs below.

Proposition D.1.5 (Non-negative deconvolution operator). Assume $\{|\xi| : \xi \in \mathscr{L}^2\} \subseteq \mathscr{L}^2(g)$, where g is a density with Fourier coefficients in $\ell^1 \cap \ell^2$. Then, $g_\bullet *$ is a non-negative operator and there exists $(g_\bullet *)^{1/2}$ such that

$$
\left\| (g_\bullet *)^{1/2} \xi_\bullet \right\|_{\ell^2}^2 = \langle g_\bullet * \xi_\bullet, \xi_\bullet \rangle_{\ell^2} = \|\xi\|_{\mathscr{L}^2(g)}^2 \, .
$$

Proof of Proposition D.1.5. Let $\xi \in \mathscr{L}^2(g)$ with $\xi = \sum_{j\in\mathbb{Z}} \xi_j e_j$, then

$$
\langle g_\bullet * \xi_\bullet, \xi_\bullet \rangle_{\ell^2} = \sum_{j\in\mathbb{Z}} \overline{\xi_j} \sum_{l\in\mathbb{Z}} g_{j-l} \xi_l = \sum_{j\in\mathbb{Z}} \overline{\xi_j} \sum_{l\in\mathbb{Z}} \mathbb{E}\left(e_l(Y) e_j(-Y) \right) \xi_l = \mathbb{E} \left| \xi(Y) \right|^2 = \|\xi\|_{\mathscr{L}^2(g)}^2 \geq 0.
$$

Hence, if for all $\xi \in \mathscr{L}^2$ we also have $|\xi| \in \mathscr{L}^2(g)$ (this is a space of real-valued functions!), then $g_\bullet *$ is a non-negative operator from ℓ^2 to ℓ^2. Consequently, there exists a non-negative operator $(g_\bullet *)^{1/2}$ such that

$$\left\|(g_\bullet *)^{1/2}\xi_\bullet\right\|_{\ell^2}^2 = \langle g_\bullet * \xi_\bullet, \xi_\bullet\rangle_{\ell^2} = \|\xi\|^2_{\mathscr{L}^2(g)}$$

for all $\xi \in \mathscr{L}^2$. □

Remark D.1.6 (Operator norm of non-negative operators). *Let $T : \mathcal{H} \to \mathcal{H}$ be a non-negative operator on the Hilbert space $\mathcal{H}$. Then there exists $T^{1/2}$ such that*

$$|\langle Th, h\rangle_{\mathcal{H}}| = \left|\langle T^{1/2}h, T^{1/2}h\rangle_{\mathcal{H}}\right| = \left\|T^{1/2}h\right\|^2_{\mathcal{H}} \qquad \forall h \in \mathcal{H}.$$

Taking the supremum over all $\|h\|_{\mathcal{H}} \leq 1$, we obtain

$$\|T\|_{\mathcal{H}\to\mathcal{H}} = \sup_{\|h\|_{\mathcal{H}}\leq 1} |\langle Th, h\rangle| = \sup_{\|h\|^2_{\mathcal{H}}\leq 1} \left\|T^{1/2}h\right\|^2_{\mathcal{H}} = \left\|T^{1/2}\right\|^2_{\mathcal{H}\to\mathcal{H}}.$$

□

D.2 Auxiliary results used in the proof of Proposition 4.2.1

Lemma D.2.1 (Control for the canonical U-statistic – indirect test). Consider $\{Y_j\}_{j=1}^n \overset{\text{iid}}{\sim} g \in \mathscr{L}^2$ and for $k \in \mathbb{N}$ the kernel $h : [0,1)^2 \longrightarrow \mathbb{R}$ given by

$$h(y_1, y_2) = \sum_{|j|\in[\![k]\!]} \frac{(e_j(-y_1) - g_j)(e_j(y_2) - \overline{g}_j)}{|\varphi_j|^2}, \qquad \forall\, y_1, y_2 \in [0,1),$$

which is real-valued, bounded, symmetric and fulfils (D.1.1). Let ν_k and m_k as in (4.2.8), then the quantities

$$\begin{aligned} A &= 4\nu_k^4 \\ B &= \sqrt{8\,\|g_\bullet\|_{\ell^2}}\nu_k^3 \leq 3\,\|g_\bullet\|_{\ell^2}\,\nu_k^3 \\ C &= D = 2\,\|g_\bullet\|_{\ell^2}\,\nu_k^2 \end{aligned} \tag{D.2.1}$$

satisfy the condition (D.1.2) in Proposition D.1.1. If, in addition, $\mathscr{L}^2(g) = \{|\xi| : \xi \in \mathscr{L}^2\}$ then also

$$D = 4\,\|g_\bullet\|_{\ell^1}\,m_k^2 \tag{D.2.2}$$

satisfies the condition (D.1.2) in Proposition D.1.1.

Proof of Lemma D.2.1. We first check the conditions on the kernel h.

1. h **is real-valued.** Let $y_1, y_2 \in [0,1)$, then

$$\overline{h(y_1, y_2)} = \sum_{|j|\in[\![k]\!]} \frac{(e_j(y_1) - \overline{g}_j)(e_j(-y_2) - g_j)}{|\varphi_j|^2}$$

where we change the summation index from j to $-l$

$$= \sum_{|l| \in [\![k]\!]} \frac{(e_{-l}(y_1) - \overline{g}_{-l})(e_{-l}(-y_2) - g_{-l})}{|\varphi_l|^2}$$

and use that $e_l(y) = e_{-l}(-y)$ for all $l \in \mathbb{Z}$

$$= \sum_{|l| \in [\![k]\!]} \frac{(e_l(-y_1) - \overline{g}_{-l})(e_l(y_2) - g_{-l})}{|\varphi_l|^2}$$

since the coefficients of (real-valued) densities satisfy $g_l = \overline{g}_{-l}$ for all $l \in \mathbb{Z}$, we have

$$= \sum_{|l| \in [\![k]\!]} \frac{(e_l(-y_1) - g_l)\,(e_l(y_2) - \overline{g}_l)}{|\varphi_l|^2} = h(y_1, y_2).$$

2. h **is symmetric.** Let $y_1, y_2 \in [0,1)$, then

$$h(y_1, y_2) = \sum_{|j| \in [\![k]\!]} \frac{(e_j(-y_1) - g_j)\left(e_j(y_2) - \overline{g}_j\right)}{|\varphi_j|^2}$$

where we change the summation index form j to $-l$

$$= \sum_{|l| \in [\![k]\!]} \frac{(e_l(y_1) - g_{-l})\,(e_l(-y_2) - \overline{g}_{-l})}{|\varphi_l|^2}$$

and use the coefficients of (real-valued) densities satisfy $g_l = \overline{g}_{-l}$ for all $l \in \mathbb{Z}$

$$= \sum_{|l| \in [\![k]\!]} \frac{(e_l(y_1) - \overline{g}_l)\,(e_l(-y_2) - g_l)}{|\varphi_l|^2} = h(y_2, y_1).$$

3. h **is bounded.** Let $y_1, y_2 \in [0,1)$, then

$$|h(y_1, y_2)| = \left| \sum_{|j| \in [\![k]\!]} \frac{(e_j(-y_1) - g_j)\left(e_j(y_2) - \overline{g}_j\right)}{|\varphi_j|^2} \right| \leq \sum_{|j| \in [\![k]\!]} \frac{\left|(e_j(-y_1) - g_j)\left(e_j(y_2) - \overline{g}_j\right)\right|}{|\varphi_j|^2}$$
$$\leq \sum_{|j| \in [\![k]\!]} \frac{4}{|\varphi_j|^2} < \infty.$$

4. h **satisfies** (D.1.1)**.** Let $Y_1 \sim g$, $y_2 \in [0,1)$, then

$$\mathbb{E}h(Y_1, y_2) = \sum_{|j| \in [\![k]\!]} \frac{(\mathbb{E}e_j(-Y_1) - g_j)\left(e_j(y_2) - \overline{g}_j\right)}{|\varphi_j|^2} = 0,$$

since $\mathbb{E}e_j(-Y_1) = g_j$.

We now first calculate quantities A, B, C that satisfy (D.1.2). Then by the discussion above, $D = C$ also satisfies (D.1.2).

1. **The quantity A.** From

$$\|(e_j - g_j)(e_{-l} - \overline{g_l})\|_{\mathscr{L}^\infty} \leq 4 \qquad \text{and } |\varphi_j| \leq 1 \qquad \text{for all } j, l \in \mathbb{Z} \tag{D.2.3}$$

we immediately obtain

$$\sup_{y_1, y_2 \in [0,1)} |h(y_1, y_2)| = \|h\|_{\mathscr{L}^\infty} \leq 4 \sum_{|j| \in [\![k]\!]} \frac{1}{|\varphi_j|^4} = 4\nu_k^4 = A.$$

2. **The quantity B.** Note that $\mathbb{E}(e_j(-Y_1)e_l(Y_1)) = \mathbb{E}(e_{j-l}(-Y_1)) = g_{j-l}$ for all $j, l \in \mathbb{Z}$. Hence, for all $y_2 \in [0,1)$ we obtain

$$\begin{aligned}
\mathbb{E}\left|h(Y_1,y_2)\right|^2 &= \operatorname{var}\left(\sum_{|j|\in[\![k]\!]} \frac{(e_j(-Y_1)-g_j)\left(e_j(y_2)-\overline{g}_j\right)}{|\varphi_j|^2}\right) \\
&= \operatorname{var}\left(\sum_{|j|\in[\![k]\!]} e_j(-Y_1)\frac{\left(e_j(y_2)-\overline{g}_j\right)}{|\varphi_j|^2}\right) \\
&\leq \mathbb{E}\left|\sum_{|j|\in[\![k]\!]} e_j(-Y_1)\frac{\left(e_j(y_2)-\overline{g}_j\right)}{|\varphi_j|^2}\right|^2 && \text{(D.2.4)} \\
&= \sum_{|j|\in[\![k]\!]}\sum_{|l|\in[\![k]\!]} \mathbb{E}(e_j(-Y_1)e_l(Y_1))\frac{\left(e_j(y_2)-\overline{g}_j\right)}{|\varphi_j|^2}\frac{(e_l(-y_2)-g_l)}{|\varphi_j|^2} \\
&= \sum_{|j|\in[\![k]\!]} \frac{\left(e_j(y_2)-\overline{g}_j\right)}{|\varphi_j|^2}\sum_{|l|\in[\![k]\!]} g_{j-l}\frac{(e_l(-y_2)-g_l)}{|\varphi_j|^2}, && \text{(D.2.5)}
\end{aligned}$$

which can be written in terms of a discrete convolution. For that purpose, let us define the $\ell^1 \cap \ell^2$-sequences $a_j := g_j \mathbb{1}_{\{|j|\in[\![2k]\!]\}}$ and $b_j := \frac{e_j(-y_2)-g_j}{|\varphi_j|^2}\mathbb{1}_{\{|j|\in[\![k]\!]\}}$ for $j \in \mathbb{Z}$. Therefore, we can write

$$\sum_{|j|\in[\![k]\!]} \frac{\left(e_j(y_2)-\overline{g}_j\right)}{|\varphi_j|^2}\sum_{|l|\in[\![k]\!]} g_{j-l}\frac{(e_l(-y_2)-g_l)}{|\varphi_j|^2} = \langle a_\bullet * b_\bullet, b_\bullet\rangle_{\ell^2}.$$

Now, we can make use of the properties of the discrete convolution operator derived in Proposition D.1.3. By 6. in Proposition D.1.3, we obtain

$$\begin{aligned}
\langle a_\bullet * b_\bullet, b_\bullet\rangle_{\ell^2} \leq \|a_\bullet\|_{\ell^1}\|b_\bullet\|_{\ell^2} &= \left(\sum_{|j|\in[\![2k]\!]}|g_j|\right)\left(\sum_{|j|\in[\![k]\!]}\frac{\left|e_j(y_2)-\overline{g}_j\right|^2}{|\varphi_j|^4}\right) \\
&\leq (4k)^{1/2}\left(\sum_{|j|\in[\![2k]\!]}|g_j|^2\right)^{1/2}\left(\sum_{|j|\in[\![k]\!]}\frac{\left|e_j(y_2)-\overline{g}_j\right|^2}{|\varphi_j|^4}\right),
\end{aligned}$$

where we applied the Cauchy-Schwarz inequality in the last step. Combining the last bound with (D.2.5), (D.2.3) and $(2k)^{1/2} \leq \nu_k^2$, it follows

$$\sup_{y_2\in[0,1)} \mathbb{E}\left|h(Y_1,y_2)\right|^2 \leq 8\nu_k^4\|g_\bullet\|_{\ell^2} = B^2.$$

3. **The quantity C.** Note that

$$\begin{aligned}
\mathbb{E}(e_j(-Y_1)-g_j)(e_l(Y_1)-\overline{g}_l) &= \mathbb{E}(e_j(-Y_1)e_l(Y_1)) - g_j\mathbb{E}e_l(Y_1) - \overline{g}_l\mathbb{E}e_j(-Y_1) + g_j\overline{g}_l \\
&= g_{j-l} - g_j\overline{g_l} - \overline{g}_l g_j + g_j\overline{g}_l = g_{j-l} - g_j\overline{g}_l
\end{aligned}$$

for all $j, l \in \mathbb{Z}$. Therefore, we obtain

$$\begin{aligned}
&\mathbb{E}\,|h(Y_1, Y_2)|^2 \\
&= \mathbb{E}\left(\sum_{|j|\in[\![k]\!]} \frac{(e_j(-Y_1)-g_j)(e_j(Y_2)-\overline{g}_j)}{|\varphi_j|^2} \sum_{|l|\in[\![k]\!]} \frac{(e_l(Y_1)-\overline{g}_l)(e_l(-Y_2)-g_l)}{|\varphi_l|^2}\right) \\
&= \sum_{|j|\in[\![k]\!]} \frac{1}{|\varphi_j|^2} \sum_{|l|\in[\![k]\!]} \frac{|g_{j-l}-g_j\overline{g}_l|^2}{|\varphi_l|^2} \\
&\le 2\sum_{|j|\in[\![k]\!]} \frac{1}{|\varphi_j|^2} \sum_{|l|\in[\![k]\!]} \frac{|g_{j-l}|^2}{|\varphi_l|^2} + 2\sum_{|j|\in[\![k]\!]} \frac{|g_j|^2}{|\varphi_j|^2} \sum_{|l|\in[\![k]\!]} \frac{|g_l|^2}{|\varphi_l|^2}. \qquad \text{(D.2.6)}
\end{aligned}$$

To the second summand we apply the Cauchy Schwarz inequality

$$2\sum_{|j|\in[\![k]\!]} \frac{|g_j|^2}{|\varphi_j|^2} \sum_{|l|\in[\![k]\!]} \frac{|g_l|^2}{|\varphi_l|^2} = 2\left(\sum_{|j|\in[\![k]\!]} \frac{|g_j|^2}{|\varphi_j|^2}\right)^2 \le 2\left(\sum_{|j|\in[\![k]\!]} |g_j|^4\right)\left(\sum_{|j|\in[\![k]\!]} \frac{1}{|\varphi_j|^4}\right) \qquad \text{(D.2.7)}$$

$$\le 2\nu_k^4 \|g_\bullet\|_{\ell^2}^2, \qquad \text{(D.2.8)}$$

where we additionally exploited that $|g_j| \le 1$ for all $j \in \mathbb{Z}$. The first summand is rewritten in terms of a discrete convolution. We define the $\ell^1 \cap \ell^2$-sequences $c_j := |g_j|^2 \mathbb{1}_{\{|j|\in[\![2k]\!]\}}$ and $d_j := \frac{1}{|\varphi_j|^2}\mathbb{1}_{\{|j|\in[\![k]\!]\}}$ for $j \in \mathbb{Z}$. Then,

$$2\sum_{|j|\in[\![k]\!]} \frac{1}{|\varphi_j|^2} \sum_{|l|\in[\![k]\!]} \frac{|g_{j-l}|^2}{|\varphi_l|^2} = 2\langle c_\bullet * d_\bullet, d_\bullet\rangle_{\ell^2}.$$

By 6. of Proposition D.1.3 this is bounded by

$$\langle c_\bullet * d_\bullet, d_\bullet\rangle_{\ell^2} \le \|c_\bullet\|_{\ell^1} \|d_\bullet\|_{\ell^2} = \|g_\bullet\|_{\ell^2}^2 \nu_k^4.$$

Combining this bound with (D.2.7) and (D.2.6), we obtain

$$\mathbb{E}\,|h(Y_1, Y_2)|^2 \le 2\|g_\bullet\|_{\ell^2}^2 \nu_k^4 + 2\|g_\bullet\|_{\ell^2} \nu_k^4 = 4\|g_\bullet\|_{\ell^2}^2 \nu_k^4 = C^2,$$

which proves the first part of the assertion.

4. **The quantity D.** Let H be the operator defined in (D.1.3) with $H\xi(y) = \mathbb{E}h(Y_1, y)\xi(Y)$, $y \in [0,1)$. Assume $\mathscr{L}^2(g) = \{|\xi| : \xi \in \mathscr{L}^2\}$. Hence, we can us the representation (D.1.6). Let $\xi \in \mathscr{L}^2(g)$, which implies $\xi = \sum_{j\in\mathbb{Z}} \xi_j e_j \in \mathscr{L}^2$. Exploiting

$$\begin{aligned}
\mathbb{E}\,(e_j(-Y_1) - g_j)\,\xi(Y_1) &= \sum_{l\in\mathbb{Z}} \xi_l \mathbb{E}e_j(-Y_1)e_l(Y_1) - g_j\mathbb{E}\xi(Y_1) = \sum_{l\in\mathbb{Z}} \xi_l g_{j-l} - g_j\mathbb{E}\xi(Y_1) \\
&= (g_\bullet * \xi_\bullet)_j - g_j\mathbb{E}\xi(Y_1)
\end{aligned}$$

and $|g_j| \le 1$ for all $j \in \mathbb{Z}$ straightforward calculations show

$$\begin{aligned}
\langle H\xi, \xi\rangle_{\mathscr{L}^2(g)} &= \int_0^1 (H\xi)(y)\xi(y)g(y)\mathrm{d}(y) = \int_0^1\int_0^1 h(z,y)\xi(z)g(z)\xi(y)g(y)\mathrm{d}z\mathrm{d}y \\
&= \sum_{|j|\in[\![k]\!]} \frac{1}{|\varphi_j|^2} \int_0^1\int_0^1 (e_j(-z)-g_j)(e_j(y)-\overline{g}_j)\xi(z)\xi(y)g(z)g(y)\mathrm{d}z\mathrm{d}y \\
&= \sum_{|j|\in[\![k]\!]} \frac{1}{|\varphi_j|^2} \left(\mathbb{E}(e_j(-Y_1)-g_j)\xi(Y_1)\right)\left(\mathbb{E}(e_j(Y_2)-\overline{g}_j)\xi(Y_2)\right) \\
&= \sum_{|j|\in[\![k]\!]} \frac{|\mathbb{E}(e_j(-Y_1)-g_j|^2}{|\varphi_j|^2} \le m_k^2 \sum_{|j|\in[\![k]\!]} |(g_\bullet * \xi_\bullet)_j - g_j\mathbb{E}\xi(Y_1)|^2 \\
&\le 2m_k^2\left(\|g_\bullet * \xi_\bullet\|_{\ell^2}^2 + \|g_\bullet\|_{\ell^1}\|\xi\|_{\mathscr{L}^2(g)}^2\right). \qquad \text{(D.2.9)}
\end{aligned}$$

Under the assumption $\mathscr{L}^2(g) = \{|\xi| : \xi \in \mathscr{L}^2\}$, the operator $g_\bullet *$ is non-negative due to Proposition D.1.5. Hence, there exists $(g_\bullet *)^{1/2}$ such that

$$(g_\bullet *)^{1/2}\left((g_\bullet *)^{1/2}\xi_\bullet\right) = g_\bullet * \xi_\bullet.$$

Since the operator norm is given by

$$\left\|(g_\bullet *)^{1/2}\right\|_{\ell^2\to\ell^2} = \sup_{\zeta_\bullet\in\ell^2,\zeta\neq 0}\frac{\left\|(g_\bullet *)^{1/2}\zeta_\bullet\right\|_{\ell^2}}{\|\zeta\|_{\ell^2}}$$

we can write (with $\zeta = (g_\bullet *)^{1/2}\xi$)

$$\begin{aligned}\|g_\bullet * \xi_\bullet\|_{\ell^2}^2 &= \left\|(g_\bullet *)^{1/2}(g_\bullet *)^{1/2}\xi_\bullet\right\|_{\ell^2}^2 \leq \left\|(g_\bullet *)^{1/2}\right\|_{\ell^2\to\ell^2}^2 \left\|(g_\bullet *)^{1/2}\xi_\bullet\right\|_{\ell^2}^2 \\ &= \|(g_\bullet *)\|_{\ell^2\to\ell^2}\|\xi\|_{\mathscr{L}^2(g)}^2 \leq \|g_\bullet\|_{\ell^1}\|\xi\|_{\mathscr{L}^2(g)}^2,\end{aligned}$$

where the second last equality is due to Proposition D.1.5 with Remark D.1.6 and the last inequality due to 6. in Proposition D.1.3. Combining this bound with (D.2.9) and the representations (D.1.5) and (D.1.6), we obtain

$$\begin{aligned}&\sup\left\{\mathbb{E}(h(Y_1,Y_2)\zeta(Y_1)\xi(Y_2), \mathbb{E}\zeta^2(Y_1)\leq 1, \mathbb{E}\xi^2(Y_2)\leq 1\right\} \\ &= \|H\|_{\mathscr{L}^2(g)\to\mathscr{L}^2(g)} = \sup\left\{\left|\langle H\xi,\xi\rangle_{\mathscr{L}^2(g)}\right| : \|\xi\|_{\mathscr{L}^2(g)}\leq 1\right\} \leq 4m_k^2\|g_\bullet\|_{\ell^1} = D.\end{aligned}$$

□

Lemma D.2.2 (Control for the linear term – indirect test). Consider random variables $\{Y_j\}_{j=1}^n \overset{\text{iid}}{\sim} g = f \circledast \varphi \in \mathscr{L}^2$ with joint distribution $\mathbb{P}_f$ and let $g^\circ = f^\circ \circledast \varphi \in \mathscr{L}^2$. For $k \in \mathbb{N}$ consider $\mathrm{q}_k^2(f - f^\circ)$ and m_k as defined in (4.2.4) and (4.2.8), respectively. Then the linear centred statistic V_n defined in (4.2.7) satisfies for all $x \geq 1$ and $n \geq 1$

$$\mathbb{P}_f\left(2\mathrm{V}_n \leq -x^2 c_1\left(1 \vee \frac{m_k^2}{n}\right)\frac{m_k^2}{n} - \frac{1}{2}\mathrm{q}_k^2(f - f^\circ)\right) \leq \exp(-x),$$

where $c_1 = 8\|g_\bullet\|_{\ell^1} + \|\varphi_\bullet\|_{\ell^2}^2$.

Proof of Lemma D.2.2. Introduce the $\mathscr{L}^2$-function $\psi := \sum_{|l|\in[\![k]\!]}\frac{g_l - g_l^\circ}{|\varphi_l|^2}e_l$ and independent and identically distributed random variables $Z_j := 2\psi(Y_j)$ for $j \in [\![n]\!]$. Note that Z_j, $j \in [\![k]\!]$ are real-valued due to the symmetry of the summation and the symmetry of the coefficients of a real-valued function, see Remark D.2.3 below. We intend to apply Proposition D.1.2 to

$$\mathrm{V}_n = \frac{1}{n}\sum_{j\in[\![n]\!]}(Z_j - \mathbb{E}_f(Z_j)).$$

For this purpose we compute the required quantities v and b. Consider b. Subsequently using the identity $g_l - g_l^\circ = (f_l - f_l^\circ)\varphi_l$ for $l \in \mathbb{Z}$, which is due to the convolution theorem, and the Cauchy-Schwarz inequality, it follows

$$\begin{aligned}|Z_1| &\leq 2\|\psi\|_{\mathscr{L}^\infty} \leq 2m_k^2\sum_{|l|\in[\![k]\!]}|g_l - g_l^\circ| = 2m_k^2\sum_{|l|\in[\![k]\!]}|(f_l - f_l^\circ)\varphi_l| \\ &\leq 2m_k^2\left(\sum_{|l|\in[\![k]\!]}|f_l - f_l^\circ|^2\right)^{1/2}\left(\sum_{|l|\in[\![k]\!]}|\varphi_l|^2\right)^{1/2} \leq 2m_k^2\mathrm{q}_k(f - f^\circ)\|\varphi_\bullet\|_{\ell^2} =: \mathrm{b}. \quad \text{(D.2.10)}\end{aligned}$$

Next, consider v. Since $\mathbb{E}_f(e_j(-Y_1)e_l(Y_1)) = g_{j-l}$ for all $j, l \in \mathbb{Z}$, we obtain

$$\begin{aligned}\mathbb{E}_f |Z_1|^2 = 4\mathbb{E}_f |\psi(Y_1)|^2 &= 4 \sum_{|j|\in[\![k]\!]} \sum_{|l|\in[\![k]\!]} \frac{\overline{g_l} - \overline{g_l^\circ}}{|\varphi_l|^2} \mathbb{E}_f\left(e_l(-Y_1)e_j(Y_1)\right) \frac{g_j - g_j^\circ}{|\varphi_j|^2} \\ &= 4 \sum_{|j|\in[\![k]\!]} \frac{\overline{g_l} - \overline{g_l^\circ}}{|\varphi_l|^2} \sum_{|l|\in[\![k]\!]} g_{j-l} \frac{g_j - g_j^\circ}{|\varphi_j|^2},\end{aligned}$$

which we rewrite in terms of a discrete convolution. Let us introduce $\ell^1 \cap \ell^2$-sequences $a_j := g_j \mathbb{1}_{\{|l|\in[\![2k]\!]\}}$ and $d_j := \frac{(g_j - g_j^\circ)}{|\varphi_j|^2} \mathbb{1}_{\{|j|\in[\![k]\!]\}}$ for $j \in \mathbb{Z}$. Exploiting 6. of Proposition D.1.3 and the identity $(g_j - g_j^\circ) = (f_j - f_j^\circ)\varphi_j$, $j \in \mathbb{Z}$, it follows

$$\begin{aligned}\mathbb{E}_f |Z_1|^2 &= 4 \sum_{|j|\in[\![k]\!]} \frac{\overline{g_l} - \overline{g_l^\circ}}{|\varphi_l|^2} \sum_{|l|\in[\![k]\!]} g_{j-l} \frac{g_j - g_j^\circ}{|\varphi_j|^2} \\ &= 4\langle a_\bullet * d_\bullet, d_\bullet \rangle \le 4 \|a_\bullet\|_{\ell^1} \|d_\bullet\|_{\ell^2}^2 \\ &\le 4 \sum_{|j|\in[\![k]\!]} \frac{\left|g_j - g_j^\circ\right|^2}{|\varphi_j|^4} \|g_\bullet\|_{\ell^1} \le 4m_k^2 \mathrm{q}_k^2(f - f^\circ) \|g_\bullet\|_{\ell^1} =: \mathrm{v}. \end{aligned} \tag{D.2.11}$$

The claim of Lemma D.2.2 now follows from Proposition D.1.2 with b and v as in (D.2.10) and (D.2.11), respectively. Indeed, making use of $2ac \le \frac{a^2}{\varepsilon} + c^2\varepsilon$ for any $a, c, \varepsilon > 0$, (D.2.10) implies for $\varepsilon_1, \varepsilon_2 > 0$

$$\frac{\mathrm{b}x}{3n} \le \varepsilon_1 \mathrm{q}_k^2(f - f^\circ) + \frac{x^2}{9\varepsilon_1} \|\varphi_\bullet\|_{\ell^2}^2 \frac{m_k^4}{n^2}$$

and

$$\sqrt{\frac{2\mathrm{v}x}{n}} \le \varepsilon_2 \mathrm{q}_k^2(f - f^\circ) + \frac{2x}{\varepsilon_2} \frac{m_k^2}{n} \|g_\bullet\|_{\ell^1}.$$

Combining both bounds with $\varepsilon_1 = \varepsilon_2 = \frac{1}{4}$ yields for all $x \ge 1$

$$\begin{aligned}\sqrt{\frac{2\mathrm{v}x}{n}} + \frac{\mathrm{b}x}{3n} &\le \frac{1}{2}\mathrm{q}_k^2(f - f^\circ) + x^2 \frac{m_k^2}{n} \left(8 \|g_\bullet\|_{\ell^1} + \frac{4}{9} \|\varphi_\bullet\|_{\ell^2}^2 \frac{m_k^2}{n}\right) \\ &\le \frac{1}{2}\mathrm{q}_k^2(f - f^\circ) + x^2 \frac{m_k^2}{n} \left(1 \vee \frac{m_k^2}{n}\right) \left(8 \|g_\bullet\|_{\ell^1} + \frac{4}{9} \|\varphi_\bullet\|_{\ell^2}^2\right),\end{aligned}$$

Hence, the assertion follows from Proposition D.1.2 by the usual symmetry argument. □

Remark D.2.3 (Z_j, $j \in [\![k]\!]$ real-valued). *Z_j, $j \in [\![k]\!]$, defined in the proof of Lemma D.2.2 are real-valued, since for all $y \in [0, 1)$, we have*

$$\overline{\psi(y)} = \overline{\sum_{|l|\in[\![k]\!]} \frac{g_l - g_l^\circ}{|\varphi_l|^2} e_l(y)} = \sum_{|l|\in[\![k]\!]} \frac{\overline{g_l} - \overline{g_l^\circ}}{|\varphi_l|^2} e_{-l}(y)$$

where we change the summation from l to $-j$

$$= \sum_{|j|\in[\![k]\!]} \frac{\overline{g_{-j}} - \overline{g_{-j}^\circ}}{|\varphi_{-j}|^2} e_j(y)$$

and exploit that for real-valued densities $g_{-j} = g_j$

$$= \sum_{|j|\in[\![k]\!]} \frac{g_j - g_j^\circ}{|\varphi_j|^2} e_j(y) = \psi(y).$$

□

D.3 Auxiliary results used in the proof of Proposition 4.4.1

Lemma D.3.1 (Control for the canonical U-statistic - direct test). Consider $\{Y_j\}_{j=1}^n \overset{\text{iid}}{\sim} g \in \mathscr{L}^2$ and for $k \in \mathbb{N}$ the kernel $h : [0,1)^2 \longrightarrow \mathbb{R}$ given by

$$h(y_1, y_2) = \sum_{|j| \in [\![k]\!]} (e_j(-y_1) - g_j)(e_j(y_2) - \overline{g}_j), \qquad \forall\, y_1, y_2 \in [0,1),$$

which is real-valued, bounded, symmetric and fulfils (D.1.1). The quantities

$$\begin{aligned} A &= 8k \\ B &= \sqrt{8\,\|g_\bullet\|_{\ell^2}}(2k)^{3/4} \le 3\,\|g_\bullet\|_{\ell^2}\,(2k)^{3/4} \\ C &= D = 2\,\|g_\bullet\|_{\ell^2}\,(2k)^{1/2} \end{aligned} \tag{D.3.1}$$

satisfy the condition (D.1.2) in Proposition D.1.1. If, in addition, $\mathscr{L}^2(g) = \{|\xi| : \xi \in \mathscr{L}^2\}$ then also

$$D = 4\,\|g_\bullet\|_{\ell^2} \tag{D.3.2}$$

satisfies the condition (D.1.2) in Proposition D.1.1.

Proof of Lemma D.3.1. Setting $|\varphi_j|^2 = 1$ for all $|j| \in [\![k]\!]$, the assertion immediately follows from Lemma D.2.1. □

Lemma D.3.2 (Control for the linear term - direct test). Let $\{Y_j\}_{j=1}^n \overset{\text{iid}}{\sim} g = f \circledast \varphi \in \mathscr{L}^2$ with joint distribution $\mathbb{P}_f$ and let $g^\circ = f^\circ \circledast \varphi \in \mathscr{L}^2$. For $k \in \mathbb{N}$ consider $\mathrm{q}_k^2(g - g^\circ)$ as defined in (4.4.1) with f, f° replaced by g, g°. Then the linear centred statistic $\mathrm{V}_n^{\mathrm{d}}$ defined in (4.4.4) satisfies for all $x \ge 1$ and $n \ge 1$

$$\mathbb{P}_f\left(2\mathrm{V}_n^{\mathrm{d}} \le -x^2 c_1 \left(1 \vee \frac{(2k)^{1/2}}{n}\right)\frac{(2k)^{1/2}}{n} - \frac{1}{2}\mathrm{q}_k^2(g - g^\circ)\right) \le \exp(-x),$$

where $c_1 = 12\,\|g_\bullet^\circ\|_{\ell^2} + 1$.

Proof of Lemma D.3.2. Introduce the $\mathscr{L}^2$-function $\psi := \sum_{|l| \in [\![k]\!]} (g_l - g_l^\circ)\, e_l$ and independent and identically distributed random variables $Z_j := 2\psi(Y_j)$ for $j \in [\![n]\!]$. Note that Z_j, $j \in [\![k]\!]$ are real-valued due to the symmetry of the coefficients of a real-valued function and the summation. We intend to apply Proposition D.1.2 to

$$\mathrm{V}_n^{\mathrm{d}} = \frac{1}{n} \sum_{j \in [\![n]\!]} (Z_j - \mathbb{E}_f(Z_j)).$$

For this purpose we compute the required quantities v and b. Consider b. By the Cauchy-Schwarz inequality, it follows

$$\begin{aligned} |Z_1| &\le 2\,\|\psi\|_{\mathscr{L}^\infty} \le 2 \sum_{|l| \in [\![k]\!]} |g_l - g_l^\circ| \le 2(2k)^{1/2} \left(\sum_{|l| \in [\![k]\!]} |g_l - g_l^\circ|^2\right)^2 \\ &= 2(2k)^{1/2}\mathrm{q}_k(g - g^\circ) =: \mathrm{b}. \end{aligned} \tag{D.3.3}$$

Next, consider v. Since $\mathbb{E}_f(e_j(-Y_1)e_l(Y_1)) = g_{j-l}$ for all $j, l \in \mathbb{Z}$, we obtain

$$\begin{aligned}\mathbb{E}_f |Z_1|^2 = 4\mathbb{E}_f |\psi(Y_1)|^2 &= 4 \sum_{|j|\in[\![k]\!]} \sum_{|l|\in[\![k]\!]} \left(\overline{g_l} - \overline{g_l^\circ}\right) \mathbb{E}_f \left(e_l(-Y_1)e_j(Y_1)\right) \left(g_j - g_j^\circ\right) \\ &= 4 \sum_{|j|\in[\![k]\!]} \left(\overline{g_l} - \overline{g_l^\circ}\right) \sum_{|l|\in[\![k]\!]} g_{j-l} \left(g_j - g_j^\circ\right),\end{aligned}$$

which we rewrite in terms of a discrete convolution. Let us introduce $\ell^1 \cap \ell^2$-sequences $a_j := g_j \mathbb{1}_{\{|l|\in[\![2k]\!]\}}$ and $d_j := (g_j - g_j^\circ)\mathbb{1}_{\{|j|\in[\![k]\!]\}}$ for $j \in \mathbb{Z}$. Exploiting 6. of Proposition D.1.3, it follows

$$\begin{aligned}\mathbb{E}_f |Z_1|^2 &= 4 \sum_{|j|\in[\![k]\!]} \left(\overline{g_l} - \overline{g_l^\circ}\right) \sum_{|l|\in[\![k]\!]} g_{j-l} \left(g_j - g_j^\circ\right) \\ &= 4\langle a_\bullet * d_\bullet, d_\bullet\rangle \leq 4 \|a_\bullet\|_{\ell^1} \|d_\bullet\|_{\ell^2}^2 \\ &\leq 4 \sum_{|j|\in[\![k]\!]} \left|g_j - g_j^\circ\right|^2 \|g_\bullet\|_{\ell^1} \leq 4\mathrm{q}_k^2(g - g^\circ) \sum_{|j|\in[\![2k]\!]} |g_j| =: \mathrm{v}. \end{aligned} \tag{D.3.4}$$

The claim of Lemma D.2.2 now follows from Proposition D.1.2 with b and v as in (D.3.3) and (D.3.4), respectively. Indeed, making use of $2ac \leq \frac{a^2}{\varepsilon} + c^2\varepsilon$ for any $a, c, \varepsilon > 0$, (D.2.10) implies for $\varepsilon_1, \varepsilon_2 > 0$

$$\frac{\mathrm{b}x}{3n} \leq \varepsilon_1 \mathrm{q}_k^2(g - g^\circ) + \frac{x^2}{9\varepsilon_1} \frac{2k}{n^2}$$

and

$$\begin{aligned}\sqrt{\frac{2\mathrm{v}x}{n}} &\leq \varepsilon_2 \mathrm{q}_k^2(f - f^\circ) + \frac{2x}{\varepsilon_2 n} \left(\sum_{|j|\in[\![2k]\!]} |g_j|\right) \\ &\leq \varepsilon_2 \mathrm{q}_k^2(f - f^\circ) + \frac{2x}{\varepsilon_2 n} (4k)^{1/2} \left(\sum_{|j|\in[\![2k]\!]} |g_j|^2\right)^{1/2} \\ &\leq \varepsilon_2 \mathrm{q}_k^2(f - f^\circ) + \frac{2\sqrt{2}x}{\varepsilon_2} \frac{(2k)^{1/2}}{n} \|g_\bullet\|_{\ell^2}.\end{aligned}$$

Combining both bounds with $\varepsilon_1 = \varepsilon_2 = \frac{1}{4}$ yields for all $x \geq 1$

$$\begin{aligned}\sqrt{\frac{2\mathrm{v}x}{n}} + \frac{\mathrm{b}x}{3n} &\leq \frac{1}{2}\mathrm{q}_k^2(g - g^\circ) + x^2 \frac{(2k)^{1/2}}{n} \left(12 \|g_\bullet\|_{\ell^2} + \frac{4}{9} \frac{(2k)^{1/2}}{n}\right) \\ &\leq \frac{1}{2}\mathrm{q}_k^2(f - f^\circ) + x^2 \frac{(2k)^{1/2}}{n} \left(1 \vee \frac{(2k)^{1/2}}{n}\right) \left(12 \|g_\bullet\|_{\ell^1} + \frac{4}{9}\right).\end{aligned}$$

Hence, the assertion follows from Proposition D.1.2 by the usual symmetry argument. □

D.4 Calculations for the χ^2-divergence

Lemma D.4.1 (χ^2-divergence over hypercubes over multiple classes). Let $\mathcal{S}$ be an arbitrary index set of finite cardinality $|\mathcal{S}| \in \mathbb{N}$. For each $s \in \mathcal{S}$ assume $k^s \in \mathbb{N}$ and $\theta_\bullet^s \in \ell^2(\mathbb{N}) \subseteq \mathbb{R}^\mathbb{N}$. For $\tau \in \{\pm\}^{k^s}$ define coefficients $\theta_\bullet^{s,\tau} \in \ell^2(\mathbb{Z})$ and functions $g^{s,\tau} \in \mathscr{L}^2$ by setting

$$\theta_j^{s,\tau} = \begin{cases} \tau_{|j|}\theta_{|j|}^s & |j| \in [\![k^s]\!] \\ 1 & j = 0 \\ 0 & |j| > k^s \end{cases} \qquad \text{and} \qquad g^{s,\tau} = \sum_{j=-k^s}^{k^s} \theta_j^{s,\tau} e_j = e_0 + \sum_{|j| \in [\![k^s]\!]} \theta_j^{s,\tau} e_j.$$

Assuming $g^{s,\tau} \in \mathcal{D}$ for each $s \in \mathcal{S}$ and $\tau \in \{\pm\}^{k^s}$, we consider the mixture $\mathbb{P}_1$ with probability density

$$\frac{1}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} \left(\frac{1}{2^{k^s}} \sum_{\tau \in \{\pm\}^{k^s}} \prod_{i \in [\![n]\!]} g^{s,\tau}(z_i) \right), \qquad \text{for } z_i \in [0,1), i \in [\![n]\!]$$

and denote $\mathbb{P}_0 = \mathbb{P}_{f^\circ}$ with probability density

$$\prod_{i \in [\![n]\!]} \mathbb{1}_{[0,1)}(z_i), \qquad \text{for } z_i \in [0,1), i \in [\![n]\!].$$

Then, the χ^2-divergence satisfies

$$\chi^2(\mathbb{P}_1, \mathbb{P}_2) \leq \frac{1}{|\mathcal{S}|^2} \sum_{s,t \in \mathcal{S}} \exp\left(2n^2 \sum_{j \in [\![k^s \wedge k^t]\!]} (\theta_j^s \theta_j^t)^2 \right) - 1.$$

Proof of Lemma D.4.1. We remind the reader of the following representation of the χ^2 divergence for measures $\mathbb{P}_1 \ll \mathbb{P}_0$ and i.i.d. random variables $(Z_j)_{j \in [\![n]\!]}$

$$\chi^2(\mathbb{P}_1, \mathbb{P}_0) = \mathbb{E}_0 \left(\frac{\mathrm{d}\mathbb{P}_1}{\mathrm{d}\mathbb{P}_0}(Z_1, \ldots, Z_n) \right)^2 - 1.$$

Let $z_j \in [0,1)$, $j \in [\![n]\!]$, then the likelihood ratio becomes

$$\frac{\mathrm{d}\mathbb{P}_1}{\mathrm{d}\mathbb{P}_0}(z_1, \ldots, z_n) = \frac{1}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} \left(\frac{1}{2^{k^s}} \sum_{\tau \in \{\pm\}^{k^s}} \prod_{i \in [\![n]\!]} g^{s,\tau}(z_i) \right),$$

since $\mathbb{P}_0$ is a product over uniform densities. Squaring, taking the expectation under $\mathbb{P}_0$ and exploiting the independence yields

$$\begin{aligned} \mathbb{E}_0 \left(\frac{\mathrm{d}\mathbb{P}_1}{\mathrm{d}\mathbb{P}_0}(Z_1, \ldots, Z_n) \right)^2 &= \frac{1}{|\mathcal{S}|^2} \sum_{s,t \in \mathcal{S}} \frac{1}{2^{k^s}} \frac{1}{2^{k^t}} \sum_{\tau \in \{\pm\}^{k^s}} \sum_{\eta \in \{\pm\}^{k^t}} \prod_{i \in [\![n]\!]} \mathbb{E}_0(g^{s,\tau}(Z_j) g^{t,\eta}(Z_j)) \\ &= \frac{1}{|\mathcal{S}|^2} \sum_{s,t \in \mathcal{S}} \frac{1}{2^{k^s}} \frac{1}{2^{k^t}} \sum_{\tau \in \{\pm\}^{k^s}} \sum_{\eta \in \{\pm\}^{k^t}} \Big(\mathbb{E}_0(g^{s,\tau}(Z_1) g^{t,\eta}(Z_1)) \Big)^n \end{aligned}$$

Let us calculate

$$\begin{aligned} \mathbb{E}_0(g^{s,\tau}(Z_1) g^{t,\eta}(Z_1)) = \int g^{s,\tau}(z) g^{t,\eta}(z) \mathrm{d}z &= \sum_{j=-k^s}^{k^s} \sum_{l=-k^t}^{k^t} \theta_j^{s,\tau} \theta_l^{t,\tau} \int e_j(z) e_l(z) \mathrm{d}z \\ &= \sum_{j=-k^s}^{k^s} \sum_{l=-k^t}^{k^t} \theta_j^{s,\tau} \theta_j^{t,\tau} \delta_{j,-l} \\ &= \sum_{j=-k^s}^{k^s} \theta_j^{s,\tau} \theta_{-j}^{t,\tau} = 1 + 2 \sum_{j \in [\![k^s \wedge k^t]\!]} \theta_j^{s,\tau} \theta_j^{t,\eta}, \end{aligned}$$

where we used the orthonormality of $(e_j)_{j\in\mathbb{Z}}$ and the symmetry of $\theta^{s,\tau}_\bullet$ respectively $\theta^{t,\eta}_\bullet$. Applying the inequality $1+x \le \exp(x)$, which holds for all $x \in \mathbb{R}$ we obtain

$$\mathbb{E}_0(g^{s,\tau}(Z_1)g^{t,\eta}(Z_1)) = 1 + 2 \sum_{j\in[\![k^s\wedge k^t]\!]} \theta_j^{s,\tau}\theta_j^{t,\eta}$$
$$\le \exp\left(2 \sum_{j\in[\![k^s\wedge k^t]\!]} \theta_j^{s,\tau}\theta_j^{t,\eta}\right) = \prod_{j\in[\![k^s\wedge k^t]\!]} \exp\left(2\theta_j^{s,\tau}\theta_j^{t,\eta}\right).$$

Hence,

$$\mathbb{E}_0\left(\frac{\mathrm{d}\mathbb{P}_1}{\mathrm{d}\mathbb{P}_0}(Z_1,\dots,Z_n)\right)^2 \le \frac{1}{|\mathcal{S}|^2}\sum_{s,t\in\mathcal{S}} \frac{1}{2^{k^s}}\frac{1}{2^{k^t}} \sum_{\tau\in\{\pm\}^{k^s}} \sum_{\eta\in\{\pm\}^{k^t}} \prod_{j\in[\![k^s\wedge k^t]\!]} \exp\left(2n\theta_j^{s,\tau}\theta_j^{t,\eta}\right),$$

where we can apply the Interchanging Lemma C.1.1 to the η-summation with $J_j^{\eta_j} = \exp(2n\theta_j^{s,\tau}\theta_j^{t,\eta})$ and obtain

$$\mathbb{E}_0\left(\frac{\mathrm{d}\mathbb{P}_1}{\mathrm{d}\mathbb{P}_0}(Z_1,\dots,Z_n)\right)^2 \le \frac{1}{|\mathcal{S}|^2}\sum_{s,t\in\mathcal{S}} \frac{1}{2^{k^s}} \sum_{\tau\in\{\pm\}^{k^s}} \prod_{j\in[\![k^s\wedge k^t]\!]} \frac{\exp\left(-2n\theta_j^{s,\tau}\theta_j^{t}\right)+\exp\left(2n\theta_j^{s,\tau}\theta_j^{t}\right)}{2}.$$

Again applying Lemma C.1.1 to the τ-summation with $J_j^{\tau_j} = \frac{\exp\left(-2n\theta_j^{s,\tau}\theta_j^{t}\right)+\exp\left(2n\theta_j^{s,\tau}\theta_j^{t}\right)}{2}$ yields

$$\mathbb{E}_0\left(\frac{\mathrm{d}\mathbb{P}_1}{\mathrm{d}\mathbb{P}_0}(Z_1,\dots,Z_n)\right)^2 \le \frac{1}{|\mathcal{S}|^2}\sum_{s,t\in\mathcal{S}} \prod_{j\in[\![k^s\wedge k^t]\!]} \frac{\exp\left(-2n\theta_j^{s}\theta_j^{t}\right)+\exp\left(2n\theta_j^{s}\theta_j^{t}\right)}{2}$$
$$= \frac{1}{|\mathcal{S}|^2}\sum_{s,t\in\mathcal{S}} \prod_{j\in[\![k^s\wedge k^t]\!]} \cosh(2n\theta_j^s\theta_j^t).$$

Since $\cosh(x) \le \exp(x^2/2)$, $x \in \mathbb{R}$ (look at the series expansions!), we obtain

$$\mathbb{E}_0\left(\frac{\mathrm{d}\mathbb{P}_1}{\mathrm{d}\mathbb{P}_0}(Z_1,\dots,Z_n)\right)^2 \le \frac{1}{|\mathcal{S}|^2}\sum_{s,t\in\mathcal{S}} \prod_{j\in[\![k^s\wedge k^t]\!]} \exp\left(2n^2\left(\theta_j^s\theta_j^t\right)^2\right)$$
$$= \frac{1}{|\mathcal{S}|^2}\sum_{s,t\in\mathcal{S}} \exp\left(2n^2 \sum_{j\in[\![k^s\wedge k^t]\!]} \left(\theta_j^s\theta_j^t\right)^2\right),$$

which completes the proof. □

Chapter 5

Testing under privacy constraints

In this chapter we investigate the compromise between protecting the privacy of an individual by transforming the data before it is released and being able to make accurate inference. We study different methods that can generate privatized versions of sensitive data that a data holder is reluctant to share. Moreover, we examine how meaningful the statistical results based on the privatized data can be. Under an additional local privacy constraints we consider the goodness-of-fit testing problem for a circular density that has already been considered in Chapter 3 and observe that standard privatization methods do not yield minimax rate, but cause a twofold deterioration of the radii.

5.1 Differential local privacy and privatized testing

Differential local privacy. In this section we assume that the *raw* sample

$$Y_k \overset{\text{iid}}{\sim} g = f \circledast \varphi, \qquad k \in [\![n]\!] \tag{5.1.1}$$

is not available to the statistician. Instead, we receive a *privatized* or *sanitized* sample $Z_k \mid Y_k = y_k$ that is obtained from $(Y_k)_{k\in[\![n]\!]}$ by a stochastic transformation $\mathbb{Q}$, called *privacy mechanism*, *stochastic channel* or *data-release mechanism*, with regular conditional distribution $\mathbb{Q}(\cdot \mid y_k)$ given $Y_k = y_k$. Formally, given two measurable spaces $(\mathcal{Y}, \sigma(\mathcal{Y}))$ and $(\mathcal{Z}, \sigma(\mathcal{Z}))$, where Y, Z take values in $\mathcal{Y}$ and $\mathcal{Z}$ respectively and are defined on a common probability space, a privacy mechanism $\mathbb{Q}$ can be associated with a Markov kernel $\kappa_{\mathbb{Q}} : (\mathcal{Y}, \sigma(\mathcal{Z})) \longrightarrow [0,1]$ with $\kappa_{\mathbb{Q}}(y_k, B) = \mathbb{P}(Z_k \in B \mid Y_k = y_k) = \mathbb{Q}(B \mid y_k)$ for all $y_k \in \mathcal{Y}$ and $B \in \sigma(Z)$. In the computer science literature, the samples $(Y_k)_{k\in[\![n]\!]}$ and $(Z_k)_{k\in[\![n]\!]}$ are often called *databases*. We assume that the stochastic channel satisfies a privacy constraint, which we formalize next. We point out that the following definition is by convention usually called α-differential local privacy. The parameter α, however, is in the context of testing also associated with the type I error, which is why we call it γ-differentially private.

Definition 5.1.1 (γ-differential privacy). Let Y be a random variables on $(\mathcal{Y}, \sigma(\mathcal{Y}))$ and let Z be random variables on $(\mathcal{Z}, \sigma(\mathcal{Z}))$, where $\sigma(\mathcal{Y})$, $\sigma(\mathcal{Z})$ are σ-fields and Y, Z are defined on a common probability space. The regular conditional distribution of Z given Y is denoted by $\mathbb{Q}$, i.e.

$$Z \mid Y = y \sim \mathbb{Q}(\cdot \mid y)$$

We call Z a **γ-differentially private view** of Y with privacy parameter $\gamma \geq 0$ if the conditional distribution satisfies

$$\mathbb{Q}(B \mid y) \leq \exp(\gamma) \cdot \mathbb{Q}(B \mid y') \qquad \text{for all } B \in \sigma(\mathcal{Z}) \text{ and } y, y' \in \mathcal{Y}. \tag{5.1.2}$$

The privacy mechanism is then called **γ-differentially locally private**. We denote the set of all γ-differentially locally private mechanism by $\mathcal{Q}_\gamma$.

The sample $(Z_k)_{k\in[\![n]\!]}$ obtained with $\mathbb{Q}$ satisfying (5.1.2) is called a **γ-differentially locally private (non-interactive)** view of the raw sample $(Y_k)_{k\in[\![n]\!]}$ (5.1.1). The term **locally** refers to the fact that to generate the kth sanitized observation Z_k we only require the kth raw observation Y_k, thus, the raw data can be stored locally. In contrast to this, there also exists the concept of **global** differential privacy, where a data collector is entrusted with the data and generates a privatized database $(Z_k)_{k\in[\![n]\!]}$ based on the entire raw data set $(Y_k)_{k\in[\![n]\!]}$. It is called **non-interactive** since we neither require the knowledge of the (possibly already generated) sanitized observations $(Z_j)_{j\leq k}$, i.e. the data holders do not need to interact with each other in order to generate the private views. Naturally, there exist many more concepts of privacy (smooth privacy, divergence-based privacy, approximate privacy etc.), for a broad overview we refer the reader to Barber and Duchi [2014].

Related literature. The concept of differential privacy was essentially introduced in the series of papers Dinur and Nissim [2003], Dwork and Nissim [2004] and Dwork [2006]. Dwork [2008] gives an overview of the early results in the field. First statistical results are derived in Wasserman and Zhou [2010] and Hall et al. [2013], where both papers work under global privacy constraints. Duchi et al. [2018] provide a toolbox of methods for deriving minimax rates of estimation under a local privacy constraint.

Let us now first heuristically explain the implications of the condition (5.1.2). A small value of γ (close to 0) corresponds to a high privacy guarantee. In the extreme case $\gamma = 0$ the conditional distributions do not depend on the value of the input data Y. Hence, we achieve total privacy. Naturally, the privatized sample is then useless for making inference on the distribution of Y. Large values of γ allow for low privacy, since a change in the original observation can then yield a completely different distribution for the output random variable and it is thus easier to draw conclusions about the raw data. Let now formalize the effect (5.1.2) has on the information about concrete input data points. Assume we want to find out whether the original (raw) data comes from Person 1 (with value y with associated probability $\mathbb{P}_0 = \mathbb{Q}(\cdot \mid y)$) or from Person 2 (with value $y' \neq y$ with associated probability $\mathbb{P}_1 = \mathbb{Q}(\cdot \mid y')$). This task can be formulated in terms of a two-point simple testing problem, which can then be solved using the Neyman-Pearson-Lemma. The privacy constraint gives a bound for the maximal power a test can achieve. The following proposition is a reformulation of Theorem 2.4. in Wasserman and Zhou [2010] and we state its proof in our setting for completeness.

Proposition 5.1.2 (Plausible deniability). Let Z be a γ-differentially private view of Y obtained through the channel $\mathbb{Q}$. Let $y \neq y'$. Any level-α-test based on the observation Z and the channel $\mathbb{Q}$ for the task

$$H_0 : \{\mathbb{P}_0 = \mathbb{Q}(\cdot \mid y)\} \qquad \text{against} \qquad H_1 : \{\mathbb{P}_1 = \mathbb{Q}(\cdot \mid y')\}$$

has power bounded by $\alpha \exp(\gamma)$.

Proof of Proposition 5.1.2. The Neyman-Pearson Lemma states that the highest possible power (i.e. minimal type II error probability) is obtained by a test of the form

$$\Delta := \mathbb{1}_{\left\{\frac{\mathrm{d}\mathbb{P}_1}{\mathrm{d}\mathbb{P}_0} \geq \tau\right\}},$$

where $\mathbb{P}_1$ and $\mathbb{P}_0$ are the probability distributions associated with the null hypothesis and the alternative, respectively, and the threshold τ satisfies

$$\mathbb{P}_0(\Delta = 1) = \mathbb{P}_0\left(\frac{\mathrm{d}\mathbb{P}_1}{\mathrm{d}\mathbb{P}_0} \geq \tau\right) \leq \alpha.$$

Note that the distributions $\mathbb{P}_0 = \mathbb{Q}(\cdot \mid y)$ and $\mathbb{P}_1 = \mathbb{Q}(\cdot \mid y')$ satisfy

$$\mathbb{P}_1(A) \leq \exp(\gamma)\mathbb{P}_0(A)$$

for any measurable set A. Hence, the power of the test is bounded by

$$1 - \mathbb{P}_1(\Delta = 0) = \mathbb{P}_1(\Delta = 1) \leq \exp(\gamma)\mathbb{P}_0(\Delta = 1) \leq \exp(\gamma)\alpha,$$

which proves the result. □

We now give two popular examples of privacy mechanisms that satisfy the privacy constraint (5.1.2). We start with a reminder of the Laplace distribution.

Reminder 5.1.3 (Laplace distribution). With $N \sim \text{Laplace}(\mu, b)$ we denote the distribution with probability density

$$f^{\text{Lp}}(x) = \frac{1}{2b}\exp\left(-\frac{|x-\mu|}{b}\right), \qquad x \in \mathbb{R}.$$

For $y, z \in \mathbb{R}$ we have $yN + z \sim \text{Laplace}(y\mu + z, yb)$. Moreover,

$$\mathbb{E}N = \mu, \qquad \operatorname{var} N = 2b^2.$$

Example 5.1.4 (Perturbation approach, "Adding noise"). The perturbation approach consists of adding centred noise N with Lebesgue density h to the observations, i.e.

$$Z := Y + N \qquad \text{with } N \sim h, \quad \mathbb{E}N = 0.$$

Then the stochastic channel $\mathbb{Q}$ has the density

$$\mathbb{q}(z \mid y) = h(z - y).$$

with respect to the Lebesgue measure. The most popular noise density is the Laplace density with appropriately chosen variance.

Example 5.1.5 (Exponential mechanism). Let $\xi : \mathcal{Y} \times \mathcal{Z} \longrightarrow [0, \infty)$ be any function and define the sensitivity of ξ by

$$\delta := \sup_{x,y \in \mathcal{Y}} \sup_{z \in Z} |\xi(x, z) - \xi(y, z)|$$

as the maximal change of ξ that can occur due to altering the input data. Define the density

$$h(z \mid y) = \frac{\exp(-\gamma \frac{\xi(y,z)}{2\delta})}{\int \exp(-\gamma \frac{\xi(y,t)}{2\delta}) \mathrm{d}t}$$

and sample $Z \sim h(\cdot \mid y)$. McSherry and Talwar [2007] show that the exponential mechanism yields a γ-differentially private channel.

Privatized testing task. We denote by $\mathcal{D}$ the probability densities in $\mathscr{L}^2$. For a nonparametric regularity class $\mathcal{E}$ and a separation set $\mathscr{L}^2_\rho := \{\xi \in \mathscr{L}^2 : \|\xi\|_{\mathscr{L}^2} \geq \rho\}$ we aim to solve the testing task

$$H_0 : f = f^\circ \qquad \text{against} \qquad H_1^\rho : f - f^\circ \in \mathscr{L}^2_\rho \cap \mathcal{E}, f \in \mathcal{D} \tag{5.1.3}$$

based on privatized views $(Z_j)_{j\in[\![n]\!]}$ of $(Y_j)_{j\in[\![n]\!]}$, where the (unobservable) raw data are independent and identically distributed copies of a circular convolution model $Y \sim g = f \circledast \varphi$. The goal is to find a pair of a privacy mechanism $\mathbb{Q}$ and a testing procedure $\{\Delta_\alpha : \alpha \in (0,1)\}$ such that (5.1.3) is solved optimally. Let us introduce a criterion for optimality. Denote by $\mathbb{P}_{f,\mathbb{Q}}$ the joint distribution of $(Z_j)_{j\in[\![n]\!]}$ if they are obtained from $Y_j \overset{\text{iid}}{\sim} g = f \circledast \varphi$, $j \in [\![n]\!]$ by applying the stochastic channel $\mathbb{Q}$. For a stochastic channel $\mathbb{Q}$, a test Δ based on the observations $(Z_j)_{j\in[\![n]\!]}$ and $\rho > 0$ we define the privatized **maximal risk** as the sum of type I and maximal type II error probabilities over the ρ-separated alternative

$$\mathcal{R}\left(\Delta, \mathbb{Q} \mid \mathcal{E}, \rho\right) := \mathbb{P}_{f^\circ,\mathbb{Q}}(\Delta(Z_1, \dots, Z_n) = 1) + \sup_{\substack{f - f^\circ \in \mathscr{L}^2_\rho \cap \mathcal{E} \\ f \in \mathcal{D}}} \mathbb{P}_{f,\mathbb{Q}}(\Delta(Z_1, \dots, Z_n) = 0).$$

The γ**-private minimax risk of testing** is then given by

$$\mathcal{R}\left(\mathcal{E}, \rho, \gamma\right) := \inf_{\mathbb{Q} \in \mathcal{Q}_\gamma} \inf_{\Delta} \mathcal{R}\left(\Delta, \mathbb{Q} \mid \mathcal{E}, \rho\right),$$

where the infimum is taken over all possible tests based on privatized observations coming from a γ-differentially private stochastic channel $\mathbb{Q}$. As usual, we search for the smallest value of ρ such that the null and the ρ-separated alternative are statistically distinguishable. A value $\rho^2 = \rho^2(\mathcal{E}, \gamma)$ is called γ**-private minimax radius of testing** if for all $\alpha \in (0,1)$ there exist constants $\underline{A}_\alpha, \overline{A}_\alpha > 0$ such that

(i) for all $A \geq \overline{A}_\alpha$ we have $\mathcal{R}\left(\mathcal{E}, \rho, \gamma\right) \leq \alpha$, (upper bound)

(ii) for all $A \leq \underline{A}_\alpha$ we have $\mathcal{R}\left(\mathcal{E}, \rho, \gamma\right) \geq 1 - \alpha$. (lower bound)

Methodology. In this chapter we fix the testing procedure (inspired by the minimax optimal procedure derived in Chapter 3) and investigate its performance in combination with different privatization methods. Throughout this chapter we assume that f and φ lie in $\mathscr{L}^2$. Our methodology heavily depends on this assumption. Consider the Fourier or exponential basis $\{e_j\}_{j\in\mathbb{Z}}$ of $\mathscr{L}^2$ with $e_j(x) := \exp(-2\pi i j x)$ for $x \in [0,1)$ and $j \in \mathbb{Z}$. Each function $\xi \in \mathscr{L}^2$ can be represented as a discrete Fourier series $\xi = \sum_{j\in\mathbb{Z}} \xi_j e_j$ where $\xi_j := \langle \xi, e_j \rangle_{\mathscr{L}^2}$ for $j \in \mathbb{Z}$. Expanding the function of interest f in the Fourier basis and applying the circular convolution theorem we obtain the representation

$$f = \sum_{j\in\mathbb{Z}} f_j e_j = \sum_{j\in\mathbb{Z}} g_j \cdot \varphi_j \cdot e_j,$$

where $(\varphi_j)_{j\in\mathbb{Z}}$ is known and $g_j = \mathbb{E}_f e_j(-Y)$, $j \in \mathbb{Z}$. Looking at the testing problem (5.1.3) it seems natural to base a test on an estimation of the quantity

$$\mathrm{q}^2(f - f^\circ) = \int_{[0,1)} (f(x) - f^\circ(x))^2 \,\mathrm{d}x = \sum_{j\in\mathbb{Z}} \left| f_j - f_j^\circ \right|^2,$$

where the last equality is due to Parseval's theorem. Using a projection approach we, in fact, estimate the truncated version

$$\mathrm{q}_k^2(f - f^\circ) = \sum_{|j|\in[\![k]\!]} \left| f_j - f_j^\circ \right|^2.$$

Note that $\mathbb{E}_f(e_j(-Y_1)) = g_j = f_j\varphi_j$, which – in case the sample $(Y_m)_{m\in[\![n]\!]}$ is available – is usually estimated by $\frac{1}{n}\sum_{m\in[\![n]\!]} e_j(-Y_m)$. This motivates why we consider privacy mechanisms generating privatized versions of the raw data $Y_m, m \in [\![n]\!]$, which are unbiased estimators of vectors $\{e_j(-Y_m)\}_{j\in[\![k]\!]}$, $m \in [\![n]\!]$ for an appropriately chosen dimension k.

Related Literature. The first result for a projection approach for estimating a density under privacy constraints is due to Wasserman and Zhou [2010], Section 6. In a non-local setting they are able to achieve the minimax rate using an orthogonal series density estimator with Laplace perturbation of the coefficients. Let us discuss the results in our model with local constraints. To our knowledge, so far there has only been work on density estimation and testing problems in direct models: Duchi et al. [2018] consider orthogonal series density estimation based on privatized views of the *direct* observations of the density;

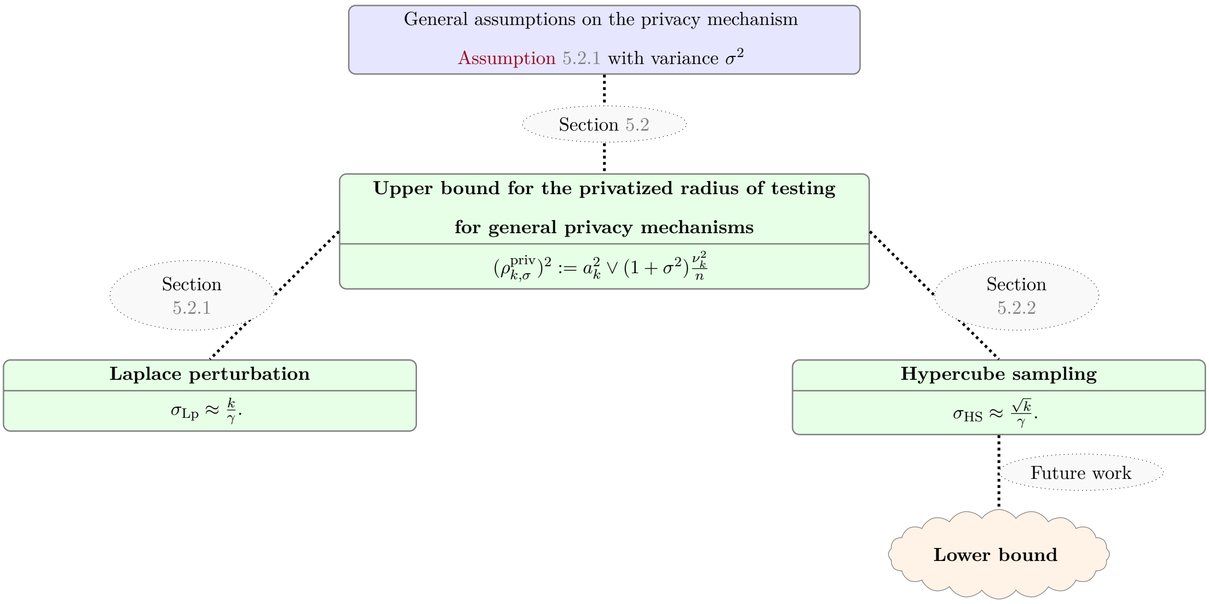
$$X_k \overset{\text{iid}}{\sim} f \in \mathscr{L}^2([0,1)), \qquad k \in [\![n]\!].$$

The non-private minimax estimation risk is well-known to be of order $n^{-\frac{2s}{2s+1}}$, where s is the smoothness parameter of a Sobolev ellipsoid (cp. (3.1.4) and Illustration 3.2.6). They show that the local γ-private minimax risk is of order $(\gamma^2 n)^{-\frac{2s}{2s+2}}$, providing both a lower and an upper bound. Butucea et al. [2020] consider Besov ellipsoids with wavelet techniques combined with a Laplace perturbation approach. Also in this case, the privatization causes a deterioration of the order of the risk from $n^{-\frac{2s}{2s+1}}$ to $(n(e^\gamma - 1)^2)^{-\frac{2s}{2s+2}}$ (s being the smoothness parameter of the Besov ellipsoid, we only state the *dense zone* here for illustration purposes). Note that this is comparable to the results of Duchi et al. [2018] since for small γ we have $\gamma \approx e^\gamma - 1$. The results mentioned so far address estimation problems. Concerning testing tasks we mention two recent papers: Lam-Weil et al. [2020] consider Besov ellipsoids with wavelet techniques combined with Laplace perturbation and show that the privatized radius of testing is sandwiched between $n^{-\frac{2s}{4s+3}}e^{-\gamma}$ (lower bound) and $(n\gamma^2)^{-\frac{2s}{4s+3}}$ (upper bound) compared to the non-private minimax radius of testing given by $n^{-\frac{2s}{4s+1}}$. Berrett and Butucea [2020] consider minimax testing of discrete distributions also for interactive mechanism. As a non-interactive mechanism they again investigate Laplace perturbations.

Naturally, also other statistical methods apart from orthogonal series approaches have been investigated under privacy constraints. Kernel estimators are, for instance, treated in Hall et al. [2013] and Kroll [2019b] under local differential approximate (γ,δ) - privacy, which is a relaxation of the constraint we consider. These papers consider Laplace perturbation and Gaussian perturbation (which is only useful in the (γ,δ)-differential privacy context). Kroll [2019b] also addresses adaptivity issues.

Outline

Testing under privacy constraints



5.2 Upper bound for general privatization methods

In this section we derive an upper bound on the γ-private minimax radius of testing under very general assumptions on the private views. Below we show that these assumptions are in particular satisfied for the Laplace-perturbation mechanism and the hypercube sampling scheme. We assume that each data holder $m \in [\![n]\!]$ releases a vector $Z_m = (Z_{m,j})_{|j| \in [\![k]\!]}$ with $Z_{m,-j} := \overline{Z_{m,j}}$ for $j \in [\![k]\!]$ containing private views of $(e_j(Y_m))_{|j| \in [\![k]\!]}$ and, thus, of Y_m. Note that the components $Z_m = (Z_{m,j})_{|j| \in [\![k]\!]}$ mimic the behaviour of $e_j(\cdot)$ and $e_{-j}(\cdot)$ and Z_m does not contain a zero element (since $e_0(Y_m) \equiv 1$ for any value of Y_m, hence, it does not need to be privatized). Therefore, data holder m only needs to generate the elements $(Z_{m,j})_{j \in [\![k]\!]}$. We denote by $\mathbb{P}_{\mathbb{Q}}(\cdot \mid Y_m)$, $\mathbb{E}_{\mathbb{Q}}(\cdot \mid Y_m)$ the distribution respectively the expectation of Z_m given Y_m. For ease of presentation from here on we only consider the case $f^\circ = \mathbb{1}_{[0,1)}$ (which translates to testing against uniformity), but note that is possible to extend the findings in this section to arbitrary $f^\circ \in \mathscr{L}^2$.

Assumption 5.2.1 (Assumptions on the private views).
For $m \in [\![n]\!]$ let $(Z_{m,j})_{j \in [\![k]\!]} \subseteq \mathbb{C}^k$ be a γ-differentially locally private view of Y_m via the channel $\mathbb{Q}$ satisfying the following four assumptions.

1. **(unbiasedness)** For all $m \in [\![n]\!]$ and $j \in [\![k]\!]$ let $\mathbb{E}_{\mathbb{Q}}(Z_{m,j} \mid Y_m) = e_j(-Y_m)$, which implies $\mathbb{E}_{f,\mathbb{Q}}(Z_{m,j}) = \mathbb{E}_f(\mathbb{E}_{\mathbb{Q}}(Z_{m,j} \mid Y_m)) = \mathbb{E}_f e_j(-Y_m) = g_j$.

2. **(independence)** For all $m, l \in [\![n]\!], m \neq l$ the vectors $(Z_{m,j})_{j \in [\![k]\!]}$ and $(Z_{l,j})_{j \in [\![k]\!]}$ are independent.

3. **(conditionally uncorrelated components)** Conditionally on Y_m the components of $(Z_{m,j})_{j \in [\![k]\!]}$ are uncorrelated, i.e. $\mathbb{E}_{\mathbb{Q}}(Z_{m,j}Z_{m,i} \mid Y_m) = \mathbb{E}_{\mathbb{Q}}(Z_{m,j} \mid Y_m)\mathbb{E}_{\mathbb{Q}}(Z_{m,i} \mid Y_m)$ almost surely for all $i, j \in [\![k]\!], i \neq j$ and $m \in [\![n]\!]$.

4. **(variance)** Conditionally on Y_m the variance is bounded by $\mathrm{var}_{\mathbb{Q}}(Z_{m,j} \mid Y_m) \leq \sigma^2$ for all $m \in [\![n]\!]$, $j \in [\![k]\!]$.

Construction of the test statistic. For $k \in \mathbb{N}$ and a privatized sample $Z_m = (Z_{m,j})_{|j| \in [\![k]\!]}$, $m \in [\![n]\!]$ we consider the quantity

$$\hat{\mathrm{p}}_k^2 := \frac{1}{n(n-1)} \sum_{\substack{m,l \in [\![n]\!] \\ l \neq m}} \sum_{|j| \in [\![k]\!]} \frac{Z_{m,j}\overline{Z_{l,j}}}{|\varphi_j|^2}, \tag{5.2.1}$$

which is an unbiased estimator of $\mathrm{q}_k^2(f - f^\circ)$ as soon as Assumption 5.2.1 (1.) is satisfied. Note that we can rewrite $\hat{\mathrm{p}}_k^2 = \frac{1}{2}\mathrm{U}_n$ with the U-statistic

$$\mathrm{U}_n := \binom{n}{2}^{-1} \sum_{\substack{m,l \in [\![n]\!] \\ l \neq m}} h(Z_m, Z_l) \tag{5.2.2}$$

with kernel $h : \mathbb{C}^{2k} \times \mathbb{C}^{2k} \longrightarrow \mathbb{R}$ given by

$$h(z_1, z_2) := \sum_{|j| \in [\![k]\!]} \frac{z_{1,j}\overline{z_{2,j}}}{|\varphi_j|^2},$$

where we index the components of a vector $z \in \mathbb{C}^{2k}$ in the following way $z = (z_j)_{|j| \in [\![k]\!]} = (z_{-k}, \ldots, z_{-1}, z_1, \ldots, z_k)$. Restricted to $V := \left\{ z \in \mathbb{C}^{2k} : z_j = \overline{z_{-j}} \right\}$ the kernel h is symmetric and real-valued. Indeed,

1. **symmetric**

$$h(z_1, z_2) = \sum_{|j| \in [\![k]\!]} \frac{z_{1,j}\overline{z_{2,j}}}{|\varphi_j|^2} = \sum_{|j| \in [\![k]\!]} \frac{\overline{z_{1,-j}}z_{2,-j}}{|\varphi_j|^2} = \sum_{|j| \in [\![k]\!]} \frac{\overline{z_{1,j}}z_{2,j}}{|\varphi_j|^2} = h(z_2, z_1),$$

2. **real-valued**

$$\overline{h(z_1, z_2)} = \sum_{|j| \in [\![k]\!]} \frac{\overline{z_{1,j}}z_{2,j}}{|\varphi_j|^2} = \sum_{|j| \in [\![k]\!]} \frac{z_{1,-j}\overline{z_{2,-j}}}{|\varphi_j|^2} = \sum_{|j| \in [\![k]\!]} \frac{z_{1,j}\overline{z_{2,j}}}{|\varphi_j|^2} = h(z_1, z_2).$$

The next proposition provides an upper bound for the variance of the estimator (5.2.1). We remark that since we assume $f, \varphi \in \mathscr{L}^2$ we have $g_\bullet = (g_j)_{j \in \mathbb{Z}} \in \ell^1(\mathbb{Z})$ by the Cauchy-Schwarz inequality. Let us also recall the notation

$$\nu_k^2 = \left(\sum_{|j| \in [\![k]\!]} \frac{1}{|\varphi_j|^4} \right)^{1/2}.$$

Proposition 5.2.2 (Upper bound for the variance). Assume that the private views satisfy Assumption 5.2.1 and consider the estimator (5.2.1). For $n \geq 2$ and any $f \in \mathscr{L}^2$ we have

$$\mathrm{var}_{f,\mathbb{Q}}(\hat{\mathrm{p}}_k^2) \leq 4 \, \|g_\bullet\|_{\ell^1} \left\{ \mathrm{q}_k^2(f - f^\circ)(1 + \sigma^2)\frac{\nu_k^2}{n} + (1 + \sigma^2)^2 \frac{\nu_k^4}{n^2} \right\}.$$

Proof of Proposition 5.2.2. Define the function $h_1 : \mathbb{C}^{2k} \longrightarrow \mathbb{C}$ by $h_1(z) := \mathbb{E}_{f,\mathbb{Q}} h(z, Z_2)$. Again, restricted to $V := \left\{ z \in \mathbb{C}^{2k} : z_j = \overline{z_{-j}} \right\}$ the function is real-valued. By Lemma A on p. 183 in Serfling [2009] the variance can be calculated by

$$\operatorname{var}_{f,\mathbb{Q}}(\mathrm{U}_n) = \binom{n}{2}^{-1} (2(n-1)\xi_1 + \xi_2) \leq \frac{4}{n}\xi_1 + \frac{4}{n^2}\xi_2$$

with

$$\xi_1 = \operatorname{var}_{f,\mathbb{Q}}(h_1(Z_1)) \qquad \text{and} \qquad \xi_2 = \operatorname{var}_{f,\mathbb{Q}}(h(Z_1, Z_2)).$$

We start with determining an upper bound for the term ξ_1. Note that due to the assumption **(unbiasedness)** in Assumption 5.2.1 we have

$$h_1(z) = \sum_{|j| \in [\![k]\!]} \frac{z_j \mathbb{E}_{f,\mathbb{Q}} \overline{Z_{2,j}}}{|\varphi_j|^2} = \sum_{|j| \in [\![k]\!]} \frac{\overline{g_j}}{|\varphi_j|^2} z_j.$$

Hence, we obtain the bound

$$\xi_1 = \operatorname{var}_{f,\mathbb{Q}}(h_1(Z_1)) \leq \mathbb{E}_{f,\mathbb{Q}} \left| \sum_{|j| \in [\![k]\!]} \frac{\overline{g_j}}{|\varphi_j|^2} Z_{1,j} \right|^2 = \sum_{|j|,|l| \in [\![k]\!]} \frac{\overline{g_j}}{|\varphi_j|^2} \frac{g_l}{|\varphi_l|^2} \mathbb{E}_{f,\mathbb{Q}}(Z_{1,j} \overline{Z_{1,l}}).$$

Let $j \notin \{\pm l\}$, then by the **conditional uncorrelatedness** and the **unbiasedness** of Assumption 5.2.1 it follows

$$\begin{aligned} \mathbb{E}_{f,\mathbb{Q}}(Z_{1,j}\overline{Z_{1,l}}) &= \mathbb{E}_f \left(\mathbb{E}_{\mathbb{Q}}(Z_{1,j}\overline{Z_{1,l}} \mid Y_1) \right) = \mathbb{E}_f \left(\mathbb{E}_{\mathbb{Q}}(Z_{1,j} \mid Y_1) \mathbb{E}_{\mathbb{Q}}(\overline{Z_{1,l}} \mid Y_1) \right) \\ &= \mathbb{E}_f(e_j(-Y_1)\overline{e_l(Y_1)}) = \mathbb{E}_f(e_{j-l}(-Y_1)) = g_{j-l}. \end{aligned}$$

Let $j \in \{\pm l\}$, then

$$\begin{aligned} \left| \mathbb{E}_{f,\mathbb{Q}}(Z_{1,j}\overline{Z_{1,l}}) \right| &\leq \mathbb{E}_{f,\mathbb{Q}} |Z_{1,j}|^2 = \operatorname{var}_{f,\mathbb{Q}}(Z_{1,j}) + |\mathbb{E}_{f,\mathbb{Q}} Z_{1,j}|^2 \\ &\leq \sigma^2 + 1 + |\mathbb{E}_f(\mathbb{E}_{\mathbb{Q}}(Z_{1,j} \mid Y_1))|^2 \\ &= \sigma^2 + 1 + |\mathbb{E}_f e_j(-Y_1)|^2 \leq \sigma^2 + 2, \end{aligned}$$

since by Eve's law we have

$$\begin{aligned} \operatorname{var}_{f,\mathbb{Q}}(Z_{1,j}) &= \mathbb{E}_f(\operatorname{var}_{\mathbb{Q}}(Z_{1,j} \mid Y_1)) + \operatorname{var}_f(\mathbb{E}_{\mathbb{Q}}(Z_{1,j} \mid Y_1)) \\ &\leq \sigma^2 + \operatorname{var}_f(e_j(-Y_1)) \leq \sigma^2 + 1. \end{aligned}$$

Hence,

$$\begin{aligned} \xi_1 &\leq \sum_{\substack{|j|,|l| \in [\![k]\!] \\ l=j}} \frac{|\overline{g_j}|}{|\varphi_j|^2} \frac{|g_l|}{|\varphi_l|^2} (\sigma^2 + 2) + \sum_{\substack{|j|,|l| \in [\![k]\!] \\ l=-j}} \frac{|\overline{g_j}|}{|\varphi_j|^2} \frac{|g_l|}{|\varphi_l|^2} (\sigma^2 + 2) + \sum_{\substack{|j|,|l| \in [\![k]\!] \\ l \notin \{\pm\}}} \frac{|\overline{g_j}|}{|\varphi_j|^2} \frac{|g_l|}{|\varphi_l|^2} |g_{j-l}| \\ &\leq 2(\sigma^2 + 2) \sum_{|j| \in [\![k]\!]} \frac{|g_j|^2}{|\varphi_j|^4} + \sum_{|j| \in [\![k]\!]} \sum_{|l| \in [\![k]\!]} \frac{|\overline{g_j}|}{|\varphi_j|^2} \frac{|g_l|}{|\varphi_l|^2} |g_{j-l}|. \end{aligned} \tag{5.2.3}$$

We define the sequences $a_\bullet$ and $b_\bullet$ by $a_j := |g_j| \mathbb{1}_{|j| \in [\![2k]\!]}$ and $b_j := \frac{|g_j|}{|\varphi_j|} \mathbb{1}_{|j| \in [\![k]\!]}$ for $j \in \mathbb{Z}$. Then, the second term can be rewritten as a discrete convolution (we refer to Section D.1, Proposition D.1.3 for the details)

$$\left| \sum_{|j| \in [\![k]\!]} \sum_{|l| \in [\![k]\!]} \frac{|\overline{g_j}|}{|\varphi_j|^2} \frac{|g_l|}{|\varphi_l|^2} |g_{j-l}| \right| = |\langle a_\bullet * b_\bullet, b_\bullet \rangle|_{\ell^2},$$

where $\langle\cdot,\cdot\rangle$ denotes the inner product of the Hilbert space $\ell^2(\mathbb{Z})$. Since $a_\bullet \in \ell^1$ and $b_\bullet \in \ell^2$ Proposition D.1.3 (6). implies

$$|\langle a_\bullet * b_\bullet, b_\bullet\rangle|_{\ell^2} \leq \|a_\bullet\|_{\ell^1} \|b_\bullet\|_{\ell^2}^2 \leq \|g_\bullet\|_{\ell^1} \sum_{|j|\in[\![k]\!]} \frac{|g_j|^2}{|\varphi_j|^4}.$$

Inserting this bound into (5.2.3) we obtain

$$\begin{aligned}\xi_1 &\leq 2(\sigma^2+2) \sum_{|j|\in[\![k]\!]} \frac{|g_j|^4}{|\varphi_j|^4} + \|g_\bullet\|_{\ell^1} \sum_{|j|\in[\![k]\!]} \frac{|g_j|^2}{|\varphi_j|^4} \\ &\leq 4(1+\sigma^2)\, \|g_\bullet\|_{\ell^1} \sum_{|j|\in[\![k]\!]} \frac{|g_j|^2}{|\varphi_j|^4}.\end{aligned}$$

Additionally applying the circular convolution theorem and the Cauchy-Schwarz inequality we obtain

$$\begin{aligned}\sum_{|j|\in[\![k]\!]} \frac{|g_j|^2}{|\varphi_j|^4} = \sum_{|j|\in[\![k]\!]} \frac{|f_j|^2}{|\varphi_j|^2} &\leq \left(\sum_{|j|\in[\![k]\!]} |f_j|^4\right)^{1/2} \left(\sum_{|j|\in[\![k]\!]} \frac{1}{|\varphi_j|^4}\right)^{1/2} \\ &\leq \sum_{|j|\in[\![k]\!]} |f_j|^2\, \nu_k^2 = \mathrm{q}_k^2(f-f^\circ)\nu_k^2,\end{aligned}$$

where we used that $\sqrt{x+y} \leq \sqrt{x} + \sqrt{y}$ for all $x, y \geq 0$ in the last line. Finally, we have the bound

$$\xi_1 \leq 4(1+\sigma^2)\, \|g_\bullet\|_{\ell^1}\, \mathrm{q}_k^2(f-f^\circ)\nu_k^2.$$

Let us now consider the term ξ_2. We have

$$\begin{aligned}\xi_2 &= \mathrm{var}_{f,\mathbb{Q}}(h(Z_1, Z_2)) \leq \mathbb{E}_{f,\mathbb{Q}}\, |h(Z_1,Z_2)|^2 \\ &= \sum_{|j|,|l|\in[\![k]\!]} \frac{1}{|\varphi_j|^2\,|\varphi_l|^2} \mathbb{E}_{f,\mathbb{Q}}\left(Z_{1,j}\overline{Z_{2,j}Z_{1,l}}Z_{2,l}\right) \\ &= \sum_{|j|,|l|\in[\![k]\!]} \frac{1}{|\varphi_j|^2\,|\varphi_l|^2} \mathbb{E}_{f,\mathbb{Q}}\left(Z_{1,j}\overline{Z_{1,l}}\right) \mathbb{E}_{f,\mathbb{Q}}\left(\overline{Z_{2,j}}Z_{2,l}\right),\end{aligned}$$

where we use the **independence** from Assumption 5.2.1. In particular for $j \in \{\pm l\}$, from the **variance** bound in Assumption 5.2.1 it follows

$$\begin{aligned}\left|\mathbb{E}_{f,\mathbb{Q}}\left(Z_{1,j}\overline{Z_{1,l}}\right)\right| &\leq \mathbb{E}_f\left(\mathbb{E}_{\mathbb{Q}}\left(\left|Z_{1,j}\overline{Z_{1,l}}\right| \mid Y_1\right)\right) \\ &= \mathbb{E}_f\left(\mathbb{E}_{\mathbb{Q}}\left(|Z_{1,j}|^2 \mid Y_1\right)\right) \\ &= \mathbb{E}_f\left(\mathrm{var}_{\mathbb{Q}}\left(Z_{1,j} \mid Y_1\right)\right) + \mathbb{E}_f\left|\mathbb{E}_{\mathbb{Q}}\left(Z_{1,j} \mid Y_1\right)\right|^2 \\ &\leq \sigma^2 + 1.\end{aligned}$$

Let $j \notin \{\pm l\}$, then due to the **conditional uncorrelatedness**

$$\begin{aligned}\mathbb{E}_{f,\mathbb{Q}}\left(Z_{1,j}\overline{Z_{1,l}}\right) &= \mathbb{E}_f\left(\mathbb{E}_{\mathbb{Q}}\left(Z_{1,j}\overline{Z_{1,l}} \mid Y_1\right)\right) = \mathbb{E}_f\left(\mathbb{E}_{\mathbb{Q}}\left(Z_{1,j} \mid Y_1\right)\mathbb{E}_{\mathbb{Q}}\left(Z_{1,l} \mid Y_1\right)\right) \\ &= \mathbb{E}_{\mathbb{Q}}\left(e_j(-Y_1)e_l(Y_1)\right) = \mathbb{E}_{\mathbb{Q}}\left(e_{j-l}(-Y_1)\right) = g_{j-l}.\end{aligned}$$

Therefore, the following bound holds

$$
\begin{aligned}
\xi_2 \leq & \sum_{\substack{|j|,|l| \in \llbracket k \rrbracket \\ j \in \{\pm l\}}} \frac{1}{|\varphi_j|^2 |\varphi_l|^2} \left|\mathbb{E}_{f,\mathbb{Q}}\left(Z_{1,j}\overline{Z_{1,l}}\right)\right| \left|\mathbb{E}_{f,\mathbb{Q}}\left(\overline{Z_{2,j}} Z_{2,l}\right)\right| \\
& + \sum_{\substack{|j|,|l| \in \llbracket k \rrbracket \\ j \notin \{\pm l\}}} \frac{1}{|\varphi_j|^2 |\varphi_l|^2} \left|\mathbb{E}_{f,\mathbb{Q}}\left(Z_{1,j}\overline{Z_{1,l}}\right)\right| \left|\mathbb{E}_{f,\mathbb{Q}}\left(\overline{Z_{2,j}} Z_{2,l}\right)\right| \\
& \leq \sum_{\substack{|j|,|l| \in \llbracket k \rrbracket \\ j \in \{\pm l\}}} \frac{1}{|\varphi_j|^2 |\varphi_l|^2} (\sigma^2+1)^2 + \sum_{\substack{|j|,|l| \in \llbracket k \rrbracket \\ j \notin \{\pm l\}}} \frac{1}{|\varphi_j|^2 |\varphi_l|^2} |g_{j-l}| \, |g_{l-j}| \\
& \leq 2(\sigma^2+1)^2 \sum_{|j| \in \llbracket k \rrbracket} \frac{1}{|\varphi_j|^4} + \sum_{|j|,|l| \in \llbracket k \rrbracket} \frac{1}{|\varphi_j|^2 |\varphi_l|^2} |g_{j-l}|^2 .
\end{aligned} \tag{5.2.4}
$$

We define the sequences $c_\bullet$ and $d_\bullet$ by $c_j := |g_j|^2 \mathbb{1}_{|j| \in \llbracket 2k \rrbracket}$ and $d_j := \frac{1}{|\varphi_j|^2} \mathbb{1}_{|j| \in \llbracket k \rrbracket}$ for $j \in \mathbb{Z}$. Then, the second term can again be written in terms of a discrete convolution. Precisely,

$$
\begin{aligned}
\left| \sum_{|j|,|l| \in \llbracket k \rrbracket} \frac{1}{|\varphi_j|^2 |\varphi_l|^2} |g_{j-l}|^2 \right| &= |\langle c_\bullet * d_\bullet, d_\bullet \rangle_{\ell^2}| \leq \|c_\bullet\|_{\ell^1} \|d_\bullet\|_{\ell^2}^2 \\
&\leq \sum_{|j| \in \llbracket 2k \rrbracket} |g_j|^4 \sum_{|j| \in \llbracket k \rrbracket} \frac{1}{|\varphi_j|^4} \leq \|g_\bullet\|_{\ell^1} \nu_k^4 .
\end{aligned}
$$

where we again applied Proposition D.1.3 (6). Inserting this bound into (5.2.4) we obtain

$$
\xi_2 \leq 2 \|g_\bullet\|_{\ell^1} (\sigma^2+1)^2 \nu_k^4 .
$$

Combining the bounds for ξ_1 and ξ_2 we get

$$
\begin{aligned}
4 \operatorname{var}_{f,\mathbb{Q}}(\hat{\mathrm{p}}^2) = \operatorname{var}_{f,\mathbb{Q}}(\mathrm{U}_n) &\leq \frac{4}{n}\xi_1 + \frac{4}{n^2}\xi_2 \\
&\leq 16(\sigma^2+1) \|g_\bullet\|_{\ell^1} \mathrm{q}_k^2(f-f^\circ) \frac{\nu_k^2}{n} + 8 \|g_\bullet\|_{\ell^1} (\sigma^2+1)^2 \frac{\nu_k^4}{n^2} \\
&\leq 16 \|g_\bullet\|_{\ell^1} \left\{ \mathrm{q}_k^2(f-f^\circ)(\sigma^2+1) \frac{\nu_k^2}{n} + (\sigma^2+1)^2 \frac{\nu_k^4}{n^2} \right\},
\end{aligned}
$$

which proves the assertion. □

Construction of the test. For $\alpha \in (0,1)$, $C_\alpha > 0$ (specified below) and $k \in \mathbb{N}$ let us consider the test

$$
\Delta_{k,\alpha}^{\mathrm{priv}} = \mathbb{1}_{\left\{ \hat{\mathrm{p}}_k^2 \geq C_\alpha (1+\sigma^2) \frac{\nu_k^2}{n} \right\}} \tag{5.2.5}
$$

based on the privatized estimator $\hat{\mathrm{p}}_k^2$ of the distance $\|f - f^\circ\|_{\mathscr{L}^2}^2$ to the null hypothesis. We consider the testing task (5.1.3) for the regularity class

$$
\mathcal{E}_{a_\bullet}^{\mathrm{R}} := \left\{ \xi \in \mathscr{L}^2 : 2 \sum_{j \in \mathbb{N}} a_j^{-2} |\xi_j|^2 \leq \mathrm{R}^2 \right\}. \tag{5.2.6}
$$

for a strictly positive, monotonically non-increasing sequence $a_\bullet = (a_j)_{j \in \mathbb{N}}$ (see Section 3.1.3 and Illustration 3.2.6 for more details about the cases covered by this general form). Furthermore, we define a privatized version of the radius of testing given by a classical bias^2-variance trade-off

$$
(\rho_{k,\sigma}^{\mathrm{priv}})^2 := a_k^2 \vee (1+\sigma^2) \frac{\nu_k^2}{n}, \tag{5.2.7}
$$

where σ^2 is the variance bound from Assumption 5.2.1 (4.). Comparing $\rho_{k,\sigma}^{\text{priv}}$ with the non-private radius ρ_k defined in (3.1.10), we see that – as expected – the privatization only has an effect on the variance-term and not on the bias^2-term. The next proposition provides an upper bound for the radius of testing of $\Delta_{k,\alpha}^{\text{priv}}$. It is an analogue result to the (non-private) upper bound derived in Corollary 3.3.3. The proof follows along similar lines as the proof of Corollary 3.3.3, using Proposition 5.2.2 instead of Corollary 3.2.5 and taking into account the private version of the threshold $(C_\alpha(1+\sigma^2)\frac{\nu_k^2}{n})$ and, therefore, appropriately modifying the case distinction. We state the proof here for completeness.

Proposition 5.2.3 (Upper bound for the radius of testing of $\Delta_{k,\alpha/2}^{\text{priv}}$).
Let $c := \|\varphi_\bullet\|_{\ell^1} < \infty$. For $\alpha \in (0,1)$ let $C_{\alpha/2}, \widetilde{A}_\alpha$ be such that

$$4 \cdot c \cdot \frac{2C_{\alpha/2}+1}{C_{\alpha/2}^2} \leq \frac{\alpha}{2} \qquad \text{and} \qquad c \cdot \frac{2C_{\alpha/2}+1}{(\widetilde{A}_\alpha - C_{\alpha/2})^2} \leq \frac{\alpha}{2} \tag{5.2.8}$$

is satisfied. Set $\overline{A}_\alpha^2 := \mathrm{R}^2 + \widetilde{A}_\alpha^2$. Let $\mathbb{Q}$ and the corresponding privatized views $(Z_1, \dots, Z_n)$ satisfy Assumption 5.2.1. Then, for all $A \geq \overline{A}_\alpha$ and all $k \in \mathbb{N}$ we obtain

$$\mathcal{R}\left(\Delta_{k,\alpha/2}^{\text{priv}}, \mathbb{Q} \mid \mathcal{E}_{a_\bullet}^{\mathrm{R}}, A\rho_{k,\sigma}^{\text{priv}}\right) \leq \alpha,$$

i.e. $(\rho_{k,\sigma}^{\text{priv}})^2$ is an upper bound for the privatized radius of testing of $\left\{\Delta_{k,\alpha/2}^{\text{priv}}\right\}_{\alpha\in(0,1)}$.

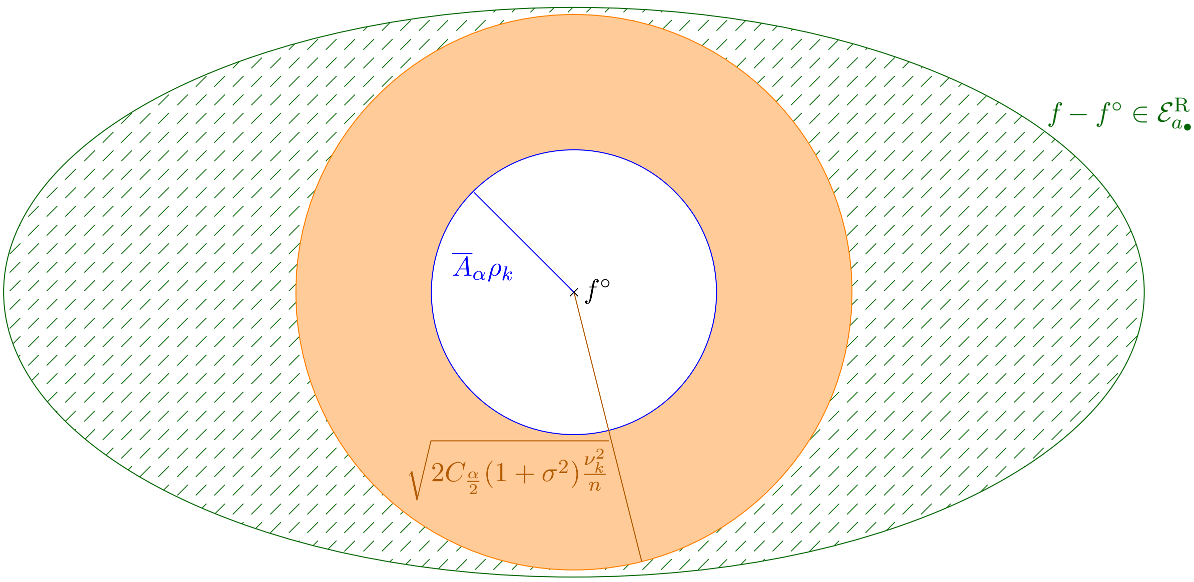


Figure 5.1: **Visualization of the structure of the proof of Proposition 5.2.3.** We distinguish the two cases: Either $f - f^\circ$ has large energy (in the first k components), hence, it is easy to test since it is far from the null (green striped area). Or $f - f^\circ$ has small energy (in the first k components), hence, it is difficult to test since it is close to the null (orange area).

Proof of Proposition 5.2.3. Recall the definition of the risk of testing

$$\mathcal{R}\left(\Delta_{k,\alpha/2}^{\text{priv}}, \mathbb{Q} \mid \mathcal{E}, A\rho_{k,\sigma}^{\text{priv}}\right) := \mathbb{P}_{f^\circ,\mathbb{Q}}(\Delta_{k,\alpha/2}^{\text{priv}} = 1) + \sup_{f - f^\circ \in \mathscr{L}^2_{\rho_{k,\sigma}^{\text{priv}}} \cap \mathcal{E}_{a_\bullet}^{\mathrm{R}}} \mathbb{P}_{f,\mathbb{Q}}(\Delta_{k,\alpha/2}^{\text{priv}} = 0).$$

We show that both the first and the second summand are bounded by $\alpha/2$. For the **type I error probability** we apply Markov's inequality and use the bound from Proposition 5.2.2. Note that in the case $f = f^\circ = \mathbb{1}_{[0,1)}$ we have $\|g_\bullet\|_{\ell^1} = 1$ and $\mathrm{q}_k^2(f - f^\circ) = 0$. Hence, we obtain

$$\begin{aligned}\mathbb{P}_{f^\circ,\mathbb{Q}}(\Delta^{\mathrm{priv}}_{k,\alpha/2} = 0) &= \mathbb{P}_{f^\circ,\mathbb{Q}}(\hat{\mathrm{p}}_k^2 \geq C_{\alpha/2}(1+\sigma^2)\frac{\nu_k^2}{n}) \\ &\leq \frac{\mathbb{E}_{f^\circ,\mathbb{Q}}\left(\hat{\mathrm{p}}_k^2\right)^2}{C^2_{\alpha/2}(1+\sigma^2)n^{-2}\nu_k^4} = \frac{\mathrm{var}_{f^\circ,\mathbb{Q}}(\hat{\mathrm{p}}_k^2)}{C^2_{\alpha/2}(1+\sigma^2)n^{-2}\nu_k^4} \leq \frac{1}{C^2_{\alpha/2}} \leq \frac{\alpha}{2}, \end{aligned} \tag{5.2.9}$$

due to (3.3.1) and $\mathbb{E}_{f^\circ,\mathbb{Q}}(\hat{\mathrm{p}}_k^2) = \mathrm{q}_k^2(f^\circ - f^\circ) = 0$. The **type II error probability** is evaluated for all f contained in the $\overline{A}_\alpha \rho^{\mathrm{priv}}_{k,\sigma}$-separated alternative, i.e. for f satisfying $f - f^\circ \in \mathcal{E}^{\mathrm{R}}_{a_\bullet}$ and $\mathrm{q}^2(f - f^\circ) \geq (\overline{A}_\alpha)^2(\rho^{\mathrm{priv}}_{k,\sigma})^2$. We centre the estimator $\hat{\mathrm{p}}_k^2$ by its expectation $\mathrm{q}_k^2(f - f^\circ)$ and obtain

$$\begin{aligned}\mathbb{P}_{f,\mathbb{Q}}(\Delta^{\mathrm{priv}}_{k,\alpha/2} = 0) &= \mathbb{P}_{f,\mathbb{Q}}(\hat{\mathrm{p}}_k^2 < C_{\alpha/2}(1+\sigma^2)\frac{\nu_k^2}{n}) \\ &= \mathbb{P}_{f,\mathbb{Q}}(\hat{\mathrm{p}}_k^2 - \mathrm{q}_k^2(f - f^\circ) < C_{\alpha/2}(1+\sigma^2)\nu_k^2 - \mathrm{q}_k^2(f - f^\circ))\end{aligned}$$

We make a case distinction on the energy of $f - f^\circ$ in the first k components.

1. $\mathrm{q}_k^2(f - f^\circ) \geq 2C_{\alpha/2}(1+\sigma^2)\frac{\nu_k^2}{n}$ (easy to test)
2. $\mathrm{q}_k^2(f - f^\circ) < 2C_{\alpha/2}(1+\sigma^2)\frac{\nu_k^2}{n}$ (difficult to test)

Case 1. (easy to test) The densities satisfying (1.) are easy to test, since the energy already contained in the first k components is large (i.e. a multiple of the standard deviation). Hence, they are easy to distinguish from the null by a test focusing on the first k components. Note that in this case we do **not** even need to use the regularity assumption $f - f^\circ \in \mathcal{E}^{\mathrm{R}}_{a_\bullet}$. Due to the case distinction we have $C_{\alpha/2}(1+\sigma^2)\frac{\nu_k^2}{n} - \mathrm{q}_k^2(f - f^\circ) \leq \frac{1}{2}\mathrm{q}_k^2(f - f^\circ)$ and, thus, Markov's inequality implies

$$\begin{aligned}\mathbb{P}_{f,\mathbb{Q}}(\Delta^{\mathrm{priv}}_{k,\alpha/2} = 0) &\leq \mathbb{P}_{f,\mathbb{Q}}(\hat{\mathrm{p}}_k^2 - \mathrm{q}_k^2(f - f^\circ) \leq -\frac{1}{2}\mathrm{q}_k^2(f - f^\circ)) \\ &= \mathbb{P}_{f,\mathbb{Q}}(\mathrm{q}_k^2(f - f^\circ) - \hat{\mathrm{p}}_k^2 \geq \frac{1}{2}\mathrm{q}_k^2(f - f^\circ)) \leq 4\frac{\mathrm{var}_{f,\mathbb{Q}}(\hat{\mathrm{p}}^2)}{(\mathrm{q}_k^2(f - f^\circ))^2}.\end{aligned}$$

Inserting the bound for the variance from Proposition 5.2.2 yields

$$\begin{aligned}\mathbb{P}_{f,\mathbb{Q}}(\Delta^{\mathrm{priv}}_{k,\alpha/2} = 0) &\leq 16\,\|g_\bullet\|_{\ell^1}\,\frac{\mathrm{q}_k^2(f - f^\circ)(1+\sigma^2)\frac{\nu_k^2}{n} + (1+\sigma^2)^2\frac{\nu_k^4}{n^2}}{(\mathrm{q}_k^2(f - f^\circ))^2} \\ &= 16\,\|g_\bullet\|_{\ell^1}\left\{\frac{(1+\sigma^2)\frac{\nu_k^2}{n}}{\mathrm{q}_k^2(f - f^\circ)} + \frac{(1+\sigma^2)^2\frac{\nu_k^4}{n^2}}{(\mathrm{q}_k^2(f - f^\circ))^2}\right\}.\end{aligned}$$

Since by the case distinction $\mathrm{q}_k^2(f - f^\circ) \geq 2C_{\alpha/2}(1+\sigma^2)\frac{\nu_k^2}{n}$, we obtain

$$\begin{aligned}\mathbb{P}_{f,\mathbb{Q}}(\Delta^{\mathrm{priv}}_{k,\alpha/2} = 0) &\leq 16\,\|g_\bullet\|_{\ell^1}\left\{\frac{1}{2C_{\alpha/2}} + \frac{1}{4C^2_{\alpha/2}}\right\} \\ &\leq \|\varphi_\bullet\|_{\ell^1}\left\{\frac{8}{C_{\alpha/2}} + \frac{4}{C^2_{\alpha/2}}\right\} \leq \alpha/2\end{aligned}$$

due to $\|g_\bullet\|_{\ell^1} \leq \|\varphi_\bullet\|_{\ell^1}$ and (5.2.8).
Case 2. (difficult to test) The densities satisfying (2.) are difficult to test, since they are

close to the detection boundary. In this case we need to exploit the separation condition of the alternative and the fact that the regularity constraint $f - f^\circ \in \mathcal{E}^{\mathrm{R}}_{a_\bullet}$ implies that, roughly speaking, the energy outside of the first k components can be controlled. In fact, for $f - f^\circ \in \mathcal{E}^{\mathrm{R}}_{a_\bullet} \cap \mathscr{L}^2_{\overline{A}_\alpha \rho^{\mathrm{priv}}_{k,\sigma}}$ we have $\sum_{|j|>k} |f - f^\circ|^2 \leq a_k^2 \mathrm{R}^2$ due to the regularity condition $f - f^\circ \in \mathcal{E}^{\mathrm{R}}_{a_\bullet}$ and $\mathrm{q}^2(f - f^\circ) = \sum_{|j|\in\mathbb{N}} |f - f^\circ|^2 \geq \overline{A}^2_\alpha (\rho^{\mathrm{priv}}_{k,\sigma})^2$ due to the energy condition $f - f^\circ \in \mathscr{L}^2_{\overline{A}_\alpha \rho^{\mathrm{priv}}_{k,\sigma}}$. Therefore,

$$\begin{aligned}
\mathrm{q}_k^2(f - f^\circ) &= \mathrm{q}^2(f - f^\circ) - \sum_{|j|>k} |f - f^\circ|^2 \geq \overline{A}^2_\alpha (\rho^{\mathrm{priv}}_{k,\sigma})^2 - a_k^2 \mathrm{R}^2 \\
&\geq \widetilde{A}^2_\alpha (1 + \sigma^2) \frac{\nu_k^2}{n} + a_k^2 \mathrm{R}^2 - a_k^2 \mathrm{R}^2 = \widetilde{A}^2_\alpha (1 + \sigma^2) \frac{\nu_k^2}{n}.
\end{aligned}$$

We use this bound and Markov's inequality to obtain

$$\begin{aligned}
\mathbb{P}_{f,\mathbb{Q}}(\Delta^{\mathrm{priv}}_{k,\alpha/2} = 0) &= \mathbb{P}_{f,\mathbb{Q}}(\hat{\mathrm{p}}_k^2 - \mathrm{q}_k^2(f - f^\circ) < C_{\alpha/2}(1 + \sigma^2)\frac{\nu_k^2}{n} - \mathrm{q}_k^2(f - f^\circ)) \\
&\leq \mathbb{P}_{f,\mathbb{Q}}(\hat{\mathrm{p}}_k^2 - \mathrm{q}_k^2(f - f^\circ) < \left(C_{\alpha/2} - \widetilde{A}^2_\alpha\right)(1 + \sigma^2)\frac{\nu_k^2}{n}) \\
&= \mathbb{P}_{f,\mathbb{Q}}(\mathrm{q}_k^2(f - f^\circ) - \hat{\mathrm{p}}_k^2 \geq \left(\widetilde{A}^2_\alpha - C_{\alpha/2}\right)(1 + \sigma^2)\frac{\nu_k^2}{n}) \\
&\leq \frac{\mathrm{var}_{f,\mathbb{Q}}(\hat{\mathrm{p}}_k^2)}{\left(\widetilde{A}^2_\alpha - C_{\alpha/2}\right)^2 (1 + \sigma^2)^2 \frac{\nu_k^4}{n^2}}.
\end{aligned}$$

Inserting the bound for the variance from Proposition 5.2.2 yields

$$\begin{aligned}
\mathbb{P}_{f,\mathbb{Q}}(\Delta^{\mathrm{priv}}_{k,\alpha/2} = 0) &\leq 4\,\|g_\bullet\|_{\ell^1} \frac{\mathrm{q}_k^2(f - f^\circ)(1 + \sigma^2)\frac{\nu_k^2}{n} + (1 + \sigma^2)^2 \frac{\nu_k^4}{n^2}}{\left(\widetilde{A}^2_\alpha - C_{\alpha/2}\right)^2 (1 + \sigma^2)^2 \frac{\nu_k^4}{n^2}} \\
&= 4\,\|g_\bullet\|_{\ell^1} \left\{ \frac{\mathrm{q}_k^2(f - f^\circ)}{\left(\widetilde{A}^2_\alpha - C_{\alpha/2}\right)^2 (1 + \sigma^2)\frac{\nu_k^2}{n}} + \frac{1}{\left(\widetilde{A}^2_\alpha - C_{\alpha/2}\right)^2} \right\} \\
&\leq 4\,\|g_\bullet\|_{\ell^1} \left\{ \frac{2C_{\alpha/2}(1 + \sigma^2)\frac{\nu_k^2}{n}}{\left(\widetilde{A}^2_\alpha - C_{\alpha/2}\right)^2 (1 + \sigma^2)\frac{\nu_k^2}{n}} + \frac{1}{\left(\widetilde{A}^2_\alpha - C_{\alpha/2}\right)^2} \right\} \\
&= 4\,\|g_\bullet\|_{\ell^1} \left\{ \frac{2C_{\alpha/2} + 1}{\left(\widetilde{A}^2_\alpha - C_{\alpha/2}\right)^2} \right\} \\
&\leq \frac{\alpha}{2},
\end{aligned}$$

where we exploited the case distinction in the third line and $\|g_\bullet\|_{\ell^1} \leq \|\varphi_\bullet\|_{\ell^1}$ and (5.2.8) in the last line, which completes the proof. □

5.2.1 Upper bound via Laplace perturbation

In this section we show that Assumption 5.2.1 is satisfied for private views generated by an appropriately applied Laplace perturbation mechanism. Thus, we can apply the general upper bound Proposition 5.2.3 to obtain an upper bound for the privatized radius of testing. Laplace perturbation is a very popular privatization mechanism and appears in the statistics literature e.g. in Wasserman and Zhou [2010], Kroll [2019b], Duchi et al. [2018], Butucea et al. [2020], Lam-Weil et al. [2020], Berrett and Butucea [2020] to mention but a few. We start with describing the perturbation method in our setting.

Description of the Laplace perturbation. Recall that the basis functions $e_j, j \in \mathbb{Z}$ are complex-valued. We perturb the real and the imaginary part separately. Let the dimension parameter $k \in \mathbb{Z}$ be fixed and let the raw data be given by $(Y_m)_{m\in[\![n]\!]}$. For $j \in [\![k]\!]$, $m \in [\![n]\!]$ and a perturbation level $b > 0$ we define

$$Z^{\mathrm{Re}}_{m,j} = \mathrm{Re}\,(e_j(Y_m)) + b\xi_{m,j},$$
$$Z^{\mathrm{Im}}_{m,j} = \mathrm{Im}\,(e_j(Y_m)) + b\zeta_{m,j}$$

with $\xi_{m,j}, \zeta_{m,j} \overset{\text{iid}}{\sim} \mathrm{Laplace}(0,1)$, i.e. due to Euler's formula $(\exp(iy) = \cos(y) + i\sin(y))$

$$Z^{\mathrm{Re}}_{m,j} \mid Y_m = y \sim \mathrm{Laplace}(\cos(2\pi jy), b),$$
$$Z^{\mathrm{Im}}_{m,j} \mid Y_m = y \sim \mathrm{Laplace}(\sin(2\pi jy), b). \tag{5.2.10}$$

Define the vector of tuples $C_m := ((Z^{\mathrm{Re}}_{m,j}, Z^{\mathrm{Im}}_{m,j}))_{j\in[\![k]\!]}$.

Proposition 5.2.4 (Privacy guarantee). $(C_m)_{m\in[\![n]\!]}$ with $b = \frac{8k}{\gamma}$ are non-interactive γ-differentially locally private views of $(Y_m)_{m\in[\![n]\!]}$.

Proof of Proposition 5.2.4. The privacy channel corresponding to the Laplace perturbation described above has density

$$\mathrm{q}(c \mid Y_m = y_m) = \prod_{|j|\in[\![k]\!]} \frac{1}{2b}\exp\left(-\frac{\left|c^{\mathrm{Re}}_j - \cos(2\pi jy_m)\right|}{b}\right)\frac{1}{2b}\exp\left(-\frac{\left|c^{\mathrm{Im}}_j - \sin(2\pi jy_m)\right|}{b}\right)$$

for $c := ((c^{\mathrm{Re}}_j, c^{\mathrm{Im}}_j))_{j\in[\![k]\!]} \in (\mathbb{R}\times\mathbb{R})^k$. For $y_m, y'_m \in [0,1)$ we consider the quotient

$$\frac{\mathrm{q}(c \mid Y_m = y_m)}{\mathrm{q}(c \mid Y_m = y'_m)}$$
$$= \prod_{|j|\in[\![k]\!]} \frac{1}{2b} e^{\frac{\left|c^{\mathrm{Re}}_j - \cos(2\pi jy'_m)\right|\left|c^{\mathrm{Re}}_j - \cos(2\pi jy_m)\right|}{b}} \frac{1}{2b} e^{\frac{\left|c^{\mathrm{Im}}_j - \sin(2\pi jy'_m)\right| - \left|c^{\mathrm{Im}}_j - \sin(2\pi jy_m)\right|}{b}}$$

Applying the reversed triangle inequality $(||a| - |b|| \le |a \pm b|)$ yields

$$\begin{aligned}\frac{\mathrm{q}(c \mid Y_m = y_m)}{\mathrm{q}(c \mid Y_m = y'_m)} &\le \prod_{|j|\in[\![k]\!]} e^{\frac{\left|\cos(2\pi jy'_m) - \cos(2\pi jy_m)\right|}{b}} e^{\frac{\left|\sin(2\pi jy'_m) - \sin(2\pi jy_m)\right|}{b}} \\ &\le \prod_{|j|\in[\![k]\!]} e^{\frac{2}{b}} e^{\frac{2}{b}} \\ &= e^{\frac{8k}{b}} \le e^{\gamma}.\end{aligned} \tag{5.2.11}$$

We have thus checked the analogous condition to (5.1.2) for densities and the proof is complete. □

Remark 5.2.5 (Order of the perturbation level b). *The proposed perturbation level $b = \frac{8k}{\gamma}$ shows the expected behaviour. For a higher level of privacy (γ small), we need to add noise with higher variance. The more evaluations of the raw data point Y_m we release, the higher the variance of the added noise should be in order to guarantee the required privacy level, i.e. b grows with k. The linear dependence of b on k is due to the Fourier basis and is specific for our situation. To be more precise it is due to the fact that all basis functions have the same support $[0, 1)$, which is reflected in the bound (5.2.11). For instance, Butucea et al. [2020] and*

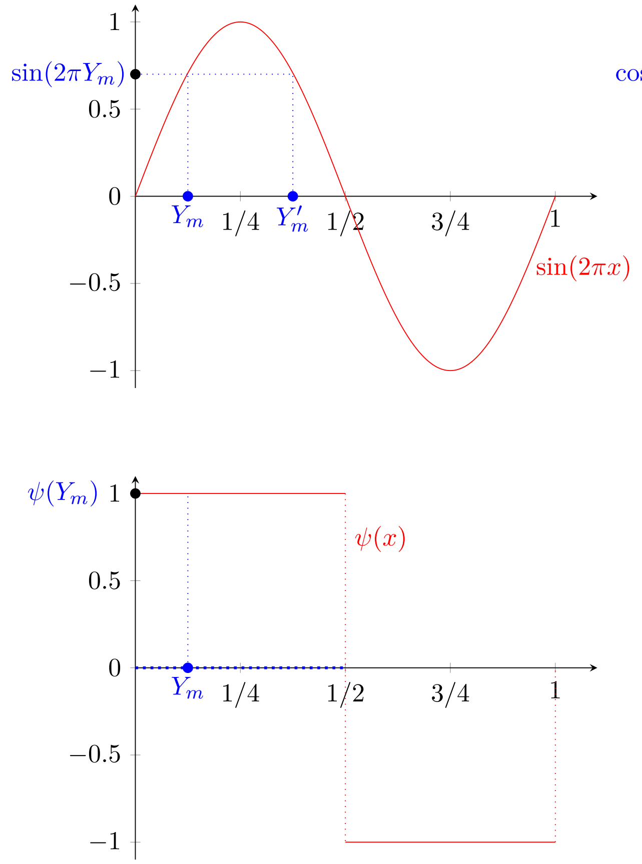


Figure 5.2: Visualization: Releasing the value $e_1(Y_m)$ respectively the tuple $(\sin(2\pi Y_m), \cos(2\pi Y_m))$ allows to uniquely identify the raw data point Y_m (first line), whereas releasing the value $\psi(Y_m)$ only provides the information $Y_m \in [0, 0.5)$ (second line). Therefore the perturbation level of the added noise that is required for γ-differentiable privacy depends on the basis.

Lam-Weil et al. [2020] consider Laplace perturbations of evaluations of wavelet bases, where the perturbation levels can be chosen of lower order (which in our situations would correspond to $\frac{\sqrt{k}}{\gamma}$*) instead. In fact, in their situations they are able to obtain a sharper bound in* (5.2.11) *due to the special structure of their wavelet bases. We point out that taking the supremum over all* $y_m, y'_m \in [0,1)$ *the bound* (5.2.11) *is indeed sharp, i.e. we cannot achieve a lower perturbation level. Let us heuristically explain why this is the case. Releasing the evaluation* $e_1(Y_m) = \cos(2\pi Y_m) + i\sin(2\pi Y_m)$ *respectively the tuple* $(\cos(2\pi Y_m), \sin(2\pi Y_m))$ *already uniquely identifies the value* Y_m*, whereas releasing the evaluation* $\psi(Y_m)$*, where* ψ *e.g. is the Haar wavelet* $\psi(x) = \mathbb{1}_{[0,1/2)}(x) - \mathbb{1}_{[1/2,1]}(x)$ *contains much less information about* Y_m*. Figure 5.2 visualizes this observation. That is to say that evaluating a finite number of Haar wavelet basis functions at a data point* Y_m *is already privatizing the data, thus, it is sufficient to add less noise.* □

Proposition 5.2.6 (Assumption 5.2.1 for Laplace perturbation). $(Z^{\mathrm{Re}}_{m,j} - iZ^{\mathrm{Im}}_{m,j})_{j \in [\![k]\!]}$, $m \in [\![n]\!]$ defined in (5.2.10) with $b = \frac{8k}{\gamma}$ are γ-differentially locally private views of Y_m, $m \in [\![n]\!]$ and satisfy Assumption 5.2.1 with $\frac{16k}{\gamma} =: \sigma_{\mathrm{LP}}$.

Proof of Proposition 5.2.6. $(Z^{\mathrm{Re}}_{m,j} - iZ^{\mathrm{Im}}_{m,j})_{j \in [\![k]\!]}$, $m \in [\![n]\!]$ are γ-differentially locally private views due to Proposition 5.2.4. We check the conditions (1.)-(4.)

1. **(unbiasedness)**

$$\mathbb{E}_{\mathbb{Q}}\left(Z^{\mathrm{Re}}_{m,j} - iZ^{\mathrm{Im}}_{m,j} \mid Y_m\right) = \mathbb{E}_{\mathbb{Q}}\left(Z^{\mathrm{Re}}_{m,j} \mid Y_m\right) - i\mathbb{E}_{\mathbb{Q}}\left(Z^{\mathrm{Im}}_{m,j} \mid Y_m\right)$$
$$= \cos(2\pi j Y_m) - i\sin(2\pi j Y_m) = e_j(-Y_m)$$

2. **(independence)** For $m \neq l$ the vectors $(Z^{\mathrm{Re}}_{m,j} - iZ^{\mathrm{Im}}_{m,j})_{j\in[\![k]\!]}$ and $(Z^{\mathrm{Re}}_{l,j} - iZ^{\mathrm{Im}}_{l,j})_{j\in[\![k]\!]}$ are independent by construction, since the Laplace perturbations conducted by data holder m and data holder l are independent.

3. **(conditionally uncorrelated components)** Conditionally on Y_m the components of $(Z^{\mathrm{Re}}_{m,j} - iZ^{\mathrm{Im}}_{m,j})_{j\in[\![k]\!]}$ are independent by construction, since the Laplace perturbations occur in each component independently.

4. **(variance)** Let $j \in [\![k]\!]$, then due to (5.2.10) and Reminder 5.1.3

$$\mathrm{var}_{\mathbb{Q}}(Z^{\mathrm{Re}}_{m,j} - iZ^{\mathrm{Im}}_{m,j}) = \mathrm{var}_{\mathbb{Q}}(Z^{\mathrm{Re}}_{m,j}) + \mathrm{var}_{\mathbb{Q}}(Z^{\mathrm{Im}}_{m,j}) = 2b^2 + 2b^2 = \left(\frac{16k}{\gamma}\right)^2 =: \sigma^2_{\mathrm{LP}}.$$

□

Inserting $\sigma = \sigma_{\mathrm{LP}} = \frac{16k}{\gamma}$ into the privatized radius of testing (5.2.7) we observe that

$$(\rho^{\mathrm{priv}}_{k,\sigma_{\mathrm{LP}}})^2 := a_k^2 \vee \left(1 + \frac{16^2k^2}{\gamma^2}\right)\frac{\nu_k^2}{n} \leq 16^2\left(\rho_k^2 \vee (\rho_k^{\mathrm{LP}})^2\right)$$

where ρ_k^2 is the non-private radius of testing (defined in (3.1.10)) and

$$(\rho_k^{\mathrm{LP}})^2 := a_k^2 \vee \frac{k^2\nu_k^2}{\gamma^2 n}.$$

The next corollary is now an immediate consequence of Proposition 5.2.3 combined with the previous Proposition 5.2.6 and we omit its proof.

Corollary 5.2.7 (Privatized radius of testing with Laplace perturbation). Let $\alpha \in (0,1)$, $\gamma \in \mathbb{R}$. Consider the family of tests $\left\{\Delta^{\mathrm{priv}}_{k,\alpha/2}\right\}, \alpha \in (0,1)$ defined in (5.2.5) and consider the privacy mechanism $\mathbb{Q}_\gamma$ associated with (5.2.10). Let $\overline{A}_\alpha$ as in Proposition 5.2.3. Then, for all $A \geq 16\overline{A}_\alpha$ and all $k \in \mathbb{N}$ we obtain

$$\mathcal{R}\left(\Delta^{\mathrm{priv}}_{k,\alpha/2}, \mathbb{Q}_\gamma \mid \mathcal{E}^{\mathrm{R}}_{a_\bullet}, A\left(\rho_k \vee \rho_k^{\mathrm{LP}}\right)\right) \leq \alpha.$$

The previous corollary shows that compared with the (non-private) radius of testing ρ_k^2 derived in Section 3.3 the privatized upper bound has the additional term $(\rho_k^{\mathrm{LP}})^2$, where the variance term is increased.

Illustration 5.2.8 (Laplace perturbation). The upper bound for the radius of testing of the tests $\Delta^{\mathrm{priv}}_{k,\alpha/2}, \alpha \in (0,1)$ and the Laplace perturbation derived in Corollary 5.2.7 depend on the dimension parameter k. Defining

$$\kappa_\star^{\mathrm{LP}} := \arg\min_{k\in\mathbb{N}}\left\{a_k^2 \vee \frac{k^2\nu_k^2}{\gamma^2 n}\right\}$$

and $\kappa_\star = \arg\min_{k\in\mathbb{N}}\left\{a_k^2 \vee \frac{\nu_k^2}{n}\right\}$ as in (3.2.5), we can optimize the upper bound with respect to

k and obtain the upper bound

$$\rho_\star^2 \vee \left(\rho_\star^{\mathrm{LP}}\right)^2 \qquad \text{with} \qquad \rho_\star^{\mathrm{LP}} := \min_{k\in\mathbb{N}} \rho_k^{\mathrm{LP}}.$$

We illustrate the order of both terms under the typical smoothness and ill-posedness assumptions introduced in Illustration 3.2.6. Compared with the (non-private) minimax radius of testing we notice two effects in the privatized term $\rho_\star^{\mathrm{LP}}$. On the one hand the sample size appears with an additional factor γ^2, which results in a smaller *effective sample size* of order $\gamma^2 n$. On the other hand in the mildly ill-posed models the radii are polynomially worse compared with the minimax radii.

Order of the upper bound for the radius of testing under Laplace perturbation			
a_j (smoothness)	$\|\varphi_j\|$ (ill-posedness)	$\rho_\star^2$	$(\rho_\star^{\mathrm{LP}})^2$
j^{-s}	$\|j\|^{-p}$	$n^{-\frac{4s}{4s+4p+1}}$	$(\gamma^2 n)^{-\frac{4s}{4s+4p+5}}$
j^{-s}	$e^{-\|j\|^p}$	$(\log n)^{-\frac{2s}{p}}$	$(\log(\gamma^2 n))^{-\frac{2s}{p}}$
e^{-j^s}	$\|j\|^{-p}$	$n^{-1}(\log n)^{\frac{4p+1}{2s}}$	$(\gamma^2 n)^{-1}(\log(\gamma^2 n))^{\frac{4p+5}{2s}}$

Calculations for the risk bounds in Illustration 5.2.8. The order of $\rho_\star^2$ has already been established in Illustration 3.2.6. Consider $(\rho_\star^{\mathrm{LP}})^2$.

1. **(ordinary smooth - mildly ill-posed)** The variance term $\frac{k^2\nu_k^2}{\gamma^2 n}$ is of order $\frac{k^{2p+5/2}}{\gamma^2 n}$ and the bias term a_k^2 is of order k^{-2s}. Hence, the optimal $\kappa_\star^{\mathrm{LP}}$ satisfies $\kappa_\star^{\mathrm{LP}} \sim (\gamma^2 n)^{\frac{2}{4s+4p+5}}$, which yields an upper bound of order $(\kappa_\star^{\mathrm{LP}})^{-2s} \sim (\gamma^2 n)^{\frac{4s}{4s+4p+5}}$.

2. **(ordinary smooth - severely ill-posed)** The variance term $\frac{k^2\nu_k^2}{\gamma^2 n}$ is of order $\frac{k^2\exp(2k^p)}{\gamma^2 n}$. Hence, the optimal $\kappa_\star^{\mathrm{LP}}$ satisfies $\kappa_\star^{\mathrm{LP}} \sim (\log(\gamma^2 n/b_{\gamma^2 n}))^{1/p}$ with $b_n \sim (\log \gamma^2 n)^{\frac{2s+2}{p}}$, which yields an upper bound of order $(\kappa_\star^{\mathrm{LP}})^{-2s} \sim (\log(\gamma^2 n))^{-\frac{2s}{p}}$.

3. **(super smooth - mildly ill-posed)** The variance term $\frac{k^2\nu_k^2}{\gamma^2 n}$ is of order $\frac{k^{2p+5/2}}{\gamma^2 n}$ and the bias term a_k^2 is of order $\exp(-2k^s)$. Hence, the optimal $\kappa_\star^{\mathrm{LP}}$ satisfies $\kappa_\star^{\mathrm{LP}} \sim (\log(\gamma^2 n/b_{\gamma^2 n}))^{1/s}$ with $b_n \sim (\log \gamma^2 n)^{\frac{4p+5}{2s}}$, which yields an upper bound of order $(\gamma^2 n)^{-1}(\log(\gamma^2 n))^{\frac{4p+5}{2s}}$.

□

Remark 5.2.9 (Naive privatization methods). *The standard technique for privatizing data is to add Laplace noise directly to the observations. Let us informally explain why this yields suboptimal results in our model. Inference on the density f of X based on observations of $Y + N$ with privatization noise N and raw data $Y \sim f \circledast \varphi$ is essentially a double-deconvolution problem. Consider for instance an ordinary smooth – mildly ill-posed model. We have already seen in Chapter 3 that a non-private testing radius cannot be of smaller order than $n^{-\frac{4s}{4s+4p+1}}$, where s is the regularity parameter and p the ill-posedness parameter of the model. Privatizing through adding (circular) noise to the observations increases the ill-posedness parameter, specifically in the situation of (wrapped) Laplace noise. Indeed, the wrapped Laplace distribution WLaplace$(0, b)$ has Fourier coefficients $f_j^{WLp} \sim (bj)^{-2}$, hence the ill-posedness parameter increases by 2 (compare Mardia and Jupp [2009], Section 3.2. for the wrapping of densities around the circumference of the circle and Comte and Taupin [2003], Section 2.1. for the coefficients). Therefore, there is no hope for obtaining a testing radius of smaller order than*

$(\gamma^2 n)^{-\frac{4s}{4s+4p+9}}$ by using this naive privatization method, which is considerably worse than what we obtained in Illustration 5.2.8. Moreover, we point out that this phenomenon is not due to the choice of the Laplace density. Naturally, there exists no density with $f_j^N \sim j^{-1/2}$ (which would yield the desired order $(\gamma^2 n)^{\frac{4s}{4s+4p+3}}$), since densities lie in $\mathscr{L}^1$. Similar observations can be made in a super smooth - mildly ill-posed model. We emphasize that in a ordinary smooth - severely ill-posed model the order of the radius of testing is already logarithmic in the sample size due to the ill-posedness of the model. In this case a privatization does not have an effect on the order (cp. the table in Illustration 5.2.8) only on the effective sample size $(\gamma^2 n)$. □

5.2.2 Upper bound via hypercube sampling

Comparing our upper bound in Corollary 5.2.7 respectively Illustration 5.2.8 with the usual upper bounds (for direct models), which are e.g. derived in Duchi et al. [2018] and Butucea et al. [2020], we see though we are able to reproduce the effect on the effective sample size $(\gamma^2 n)$ we have an additional deterioration in the exponent (where a 3 instead of the 5 in both the (ordinary smooth - mildly ill-posed) and (super smooth - mildly ill-posed) case appears). This deterioration directly translates to the higher perturbation level that we remarked on in Remark 5.2.5. Therefore, we consider another privatization mechanism, which is less standard but better suited for our model. Heuristically, the hypercube sampling mechanism described below works better for our choice of basis (which is determined by our testing approach) than the Laplace perturbation, since it samples all components of the privatized views simultaneously instead of adding noise separately to each component. The hypercube sampling mechanism was introduced in Duchi et al. [2018], p.17 and we adapt it here to our setting.

Description of the hypercube sampling mechanism. We describe the procedure of data holder m with given raw data $Y_m = Y = y$ and omit the index m for readability.

1. Step (Evaluation): Create the vectors
$$v = (v_j)_{j\in[\![k]\!]} = (\cos(2\pi jy))_{j\in[\![k]\!]} \qquad w = (w_j)_{j\in[\![k]\!]} = (\sin(2\pi jy))_{j\in[\![k]\!]}$$
by evaluating the k basis functions e_j, $j \in [\![k]\!]$ at $Y = y$ and storing the real part in v and the imaginary part in w. Denote by $V = (\mathrm{Re}\, e_j(Y))_{j\in[\![k]\!]}$ and $W = (\mathrm{Im}\, e_j(Y))_{j\in[\![k]\!]}$ the corresponding random variables.

2. Step (Cube Sampling): Sample vertices of a k-hypercube $\tilde{Y}^{\mathrm{Re}} = (\tilde{Y}_j^{\mathrm{Re}})_{j\in[\![k]\!]}$ and $\tilde{Y}^{\mathrm{Im}} = (\tilde{Y}_j^{\mathrm{Im}})_{j\in[\![k]\!]}$ with independent components according to
$$\mathbb{P}(\tilde{Y}_j^{\mathrm{Re}} = 1 \mid V = v) = \frac{1}{2} + \frac{v_j}{2}, \qquad \mathbb{P}(\tilde{Y}_j^{\mathrm{Re}} = -1 \mid V = v) = \frac{1}{2} - \frac{v_j}{2}$$
$$\mathbb{P}(\tilde{Y}_j^{\mathrm{Im}} = 1 \mid W = w) = \frac{1}{2} + \frac{w_j}{2}, \qquad \mathbb{P}(\tilde{Y}_j^{\mathrm{Im}} = -1 \mid W = w) = \frac{1}{2} - \frac{w_j}{2}$$

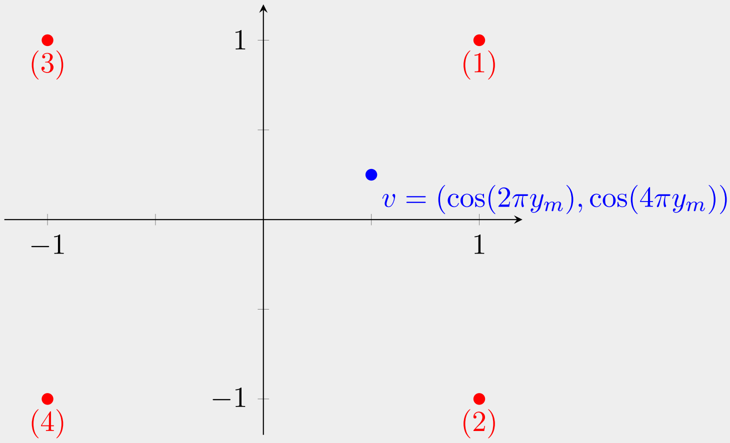
3. Step (Privatization) Sample the Bernoulli random variables T^{Re}, T^{Im} independently according to
$$\mathbb{P}(T^{\mathrm{Re}} = 0) = \mathbb{P}(T^{\mathrm{Im}} = 0) = \frac{1}{e^{\gamma/2}+1}, \quad \mathbb{P}(T^{\mathrm{Re}} = 1) = \mathbb{P}(T^{\mathrm{Im}} = 1) = \frac{e^{\gamma/2}}{e^{\gamma/2}+1}.$$

4. Step (Cube Sampling): Sample vertices of a k-hypercube Z^{Im}, Z^{Re} according to
$$Z^{\mathrm{Re}} \sim \begin{cases} U(z \in \{\pm B\}^k \mid \langle z, \tilde{Y}^{\mathrm{Re}} \rangle \geq 0) & \text{if } T^{\mathrm{Re}} = 1, \\ U(z \in \{\pm B\}^k \mid \langle z, \tilde{Y}^{\mathrm{Re}} \rangle \leq 0) & \text{if } T^{\mathrm{Re}} = 0 \end{cases}$$
$$Z^{\mathrm{Im}} \sim \begin{cases} U(z \in \{\pm B\}^k \mid \langle z, \tilde{Y}^{\mathrm{Im}} \rangle \geq 0) & \text{if } T^{\mathrm{Im}} = 1, \\ U(z \in \{\pm B\}^k \mid \langle z, \tilde{Y}^{\mathrm{Im}} \rangle \leq 0) & \text{if } T^{\mathrm{Im}} = 0 \end{cases}$$

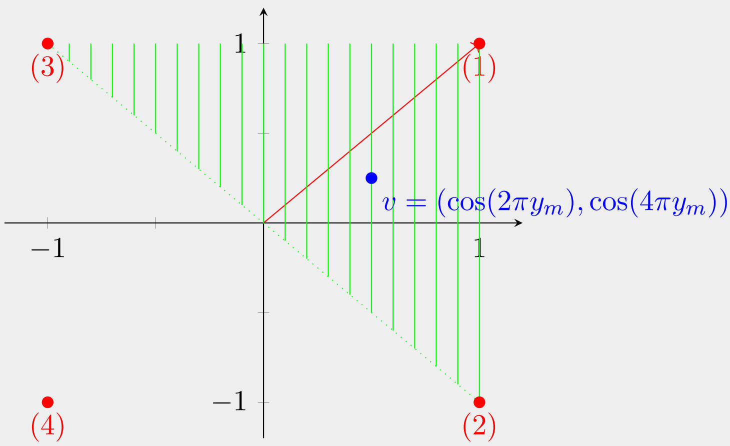
where U denotes the uniform distribution on a discrete set, $B := \frac{1}{c_k}\frac{e^{\gamma/2}+1}{e^{\gamma/2}-1}$ and

$$c_k := \begin{cases} \frac{1}{2^{k-1}}\binom{k-1}{(k-1)/2} & k \text{ odd,} \\ \frac{1}{2^{k-1}+1/2\binom{k}{k/2}}\binom{k-1}{k/2} & k \text{ even.} \end{cases}$$

Illustration 5.2.10. We illustrate the hypercube sampling scheme for the real part and $k = 2$. Assume that $Y_m = y_m$ is the given raw data point. Let $v = (\cos(2\pi y_m), \cos(4\pi y_m)) = (\text{Re }(e_1(y_m)), \text{Re }(e_2(y_m))$, i.e. the real parts of the evaluations of the first two basis functions. We describe the procedure to obtain a privatized view of v via the hypercube sampling mechanism.



Projection onto cube: We plot the vector v of the real part of the evaluations of the first two basis functions. The vertices of the hypercube (red points) in which this vector lies are assigned probabilities. The closer a vertex is to the raw data point v, the higher the probability. The vertices in the picture are numbered with decreasing probability, i.e. (1) is the most likely to be sampled in the first step.



Privatization and sampling step: Assume that indeed vertex $\tilde{Y}_m = (1)$ was sampled in the first step. This defines a hyperplane, the striped part (green) is given by $\left\{z \in \mathbb{R}^2 : \langle z, \tilde{Y}_m\rangle \geq 0\right\}$. Sampling $T \sim \text{Ber}(\frac{e^\gamma}{e^\gamma+1})$ decides whether we sample uniformly from the vertices of the hypercube $\{\pm B\}^k$ in the *correct* (green) hyperplane (in case of a *success* $T = 1$) or from the hyperplane that contains vertices further apart from the raw data point (in case of a *failure* $T = 0$).

By Duchi et al. [2018] (p.17) Z^{Im} and Z^{Re} are $\gamma/2$-differentially private views of Y. We want to combine the two into one private view via the following well-known lemma (see e.g. Kroll [2019b], Lemma 2.16. for a proof)

Lemma 5.2.11 (Composition lemma). Let Z_1, Z_2 be γ_1 respectively γ_2-differentially locally private views of Y, which are independent conditionally on Y. Then (Z_1, Z_2) is a $(\gamma_1 + \gamma_2)$-differentially locally private view of Y.

Hence, Lemma 5.2.11 implies that $Z^{\mathrm{Re}} - iZ^{\mathrm{Im}}$ is a γ-differentially private view of Y. Assume now that each data holder m carries out the mechanism to generate views $Z_m = (Z^{\mathrm{Re}}_{m,j} - iZ^{\mathrm{Im}}_{m,j})_{j\in[\![k]\!]}$.

Proposition 5.2.12 (Assumption 5.2.1 for hypercube sampling). $(Z^{\mathrm{Re}}_{m,j} - iZ^{\mathrm{Im}}_{m,j})_{j\in[\![k]\!]}$, $m \in [\![n]\!]$ sampled according to the hypercube mechanism are γ-differentially locally private views of Y_m, $m \in [\![n]\!]$ and satisfy Assumption 5.2.1 with $c \cdot \sqrt{k}\frac{e^{\gamma/2}+1}{e^{\gamma/2}-1} =: \sigma_{\mathrm{HS}}$ for a universal constant $c > 0$.

Proof of Proposition 5.2.12.

1. **(unbiasedness)**

$$\mathbb{E}_{\mathbb{Q}}\left(Z^{\mathrm{Re}}_{m,j} - iZ^{\mathrm{Im}}_{m,j} \mid Y_m\right) = \mathbb{E}_{\mathbb{Q}}\left(Z^{\mathrm{Re}}_{m,j} \mid Y_m\right) - i\mathbb{E}_{\mathbb{Q}}\left(Z^{\mathrm{Im}}_{m,j} \mid Y_m\right)$$
$$= \cos(2\pi j Y_m) - i\sin(2\pi j Y_m) = e_j(-Y_m),$$

which follows from $I.3$ in the appendix of Duchi et al. [2018].

2. **(independence)** For $m \neq l$ the vectors $(Z^{\mathrm{Re}}_{m,j} - iZ^{\mathrm{Im}}_{m,j})_{j\in[\![k]\!]}$ and $(Z^{\mathrm{Re}}_{l,j} - iZ^{\mathrm{Im}}_{l,j})_{j\in[\![k]\!]}$ are independent by construction, since the sampling schemes conducted by data holder m and data holder l are independent.

3. **(conditionally uncorrelated components)** Conditionally on Y_m the components of $(Z^{\mathrm{Re}}_{m,j} - iZ^{\mathrm{Im}}_{m,j})_{j\in[\![k]\!]}$ are independent by construction, since the components of $\tilde{Y}^{\mathrm{Re}}$ and $\tilde{Y}^{\mathrm{Im}}$ are sampled independently.

4. **(variance)** Let $j \in [\![k]\!]$, then due to (5.2.10) and Reminder 5.1.3

$$\mathrm{var}_{\mathbb{Q}}(Z^{\mathrm{Re}}_{m,j} - iZ^{\mathrm{Im}}_{m,j}) = \mathrm{var}_{\mathbb{Q}}(Z^{\mathrm{Re}}_{m,j}) + \mathrm{var}_{\mathbb{Q}}(Z^{\mathrm{Im}}_{m,j}) \leq B^2 + B^2$$
$$\leq \left(c \cdot \sqrt{k}\frac{e^{\gamma/2}+1}{e^{\gamma/2}-1}\right)^2 =: \sigma^2_{\mathrm{HS}},$$

due to the fact that the Z^{Im}_m and Z^{Re}_m lie in a bounded cube and Stirling's approximation. Indeed, applying Stirling's formula to all three factorials shows that for any $n, 2 \leq k \leq n$

$$\binom{n}{k} = (1 + o(1))\sqrt{\frac{n}{2\pi k(n-k)}}\left(\frac{n}{k}\right)^k\left(\frac{n}{n-k}\right)^{n-k}.$$

Let us first consider the case when k is odd, then

$$c_k = \frac{1}{2^{k-1}}\binom{k-1}{(k-1)/2} = (1 + o(1))\,\frac{1}{2^{k-1}}\sqrt{\frac{2}{\pi(k-1)}}2^{k-1} = (1 + o(1))$$
$$\leq (1 + o(1))\,\sqrt{\frac{4}{\pi k}} \leq \frac{1}{c}\frac{1}{\sqrt{k}}.$$

Now let k be even, then

$$\begin{aligned}
c_k &= \frac{1}{2^{k-1} + 1/2\binom{k}{k/2}} \binom{k-1}{k/2} \\
&\leq (1+o(1)) \frac{1}{2^{k-1}} \sqrt{\frac{k-1}{2\pi k/2(k/2-1)}} \left(\frac{k-1}{k/2}\right)^{k/2} \left(\frac{k-1}{k/2-1}\right)^{k/2-1} \\
&\leq (1+o(1)) \frac{1}{2^{k-1}} \sqrt{\frac{2}{\pi k}} 2^{k-1} \\
&\leq \frac{1}{c} \frac{1}{\sqrt{k}}.
\end{aligned}$$

□

Corollary 5.2.13 (Assumption 5.2.1 for hypercube sampling). Let $\gamma \in (0,1)$. $(Z^{\mathrm{Re}}_{m,j} - iZ^{\mathrm{Im}}_{m,j})_{j \in [\![k]\!]}$, $m \in [\![n]\!]$ sampled according to the hypercube mechanism are γ-differentially locally private views of Y_m, $m \in [\![n]\!]$ and satisfy Assumption 5.2.1 with $c \cdot \frac{\sqrt{k}}{\gamma} = \sigma$ for a universal constant $c > 0$.

Proof. Note that $1 + x \leq e^x$ for all $x \in \mathbb{R}$. Thus,

$$\frac{e^{\gamma/2}+1}{e^{\gamma/2}-1} = \frac{e^{\gamma/2}-1}{e^{\gamma/2}-1} + \frac{2}{e^{\gamma/2}-1} \leq 1 + \frac{2}{\gamma/2} \leq \frac{5}{\gamma},$$

then the assertion follows from Proposition 5.2.12. □

Comparing the variance $c \cdot \frac{\sqrt{k}}{\gamma} = \sigma$ of the private views obtained via hypercube sampling with the private views from the Laplace perturbation (cp. Proposition 5.2.6), we see that we have improved by a factor of $\sqrt{k}$. Inserting $\sigma = \frac{c\sqrt{k}}{\gamma}$ into the privatized radius of testing (5.2.7) we observe that

$$(\rho^{\mathrm{priv}}_{k,\sigma_{\mathrm{HS}}})^2 := a_k^2 \vee \left(1 + \frac{c^2 k}{\gamma^2}\right) \frac{\nu_k^2}{n} \leq \rho_k^2 \vee (\rho_k^{\mathrm{HS}})^2,$$

where ρ_k^2 is the non-private radius of testing (defined in (3.1.10)) and

$$(\rho_k^{\mathrm{HS}})^2 := a_k^2 \vee c^2 \frac{k\nu_k^2}{\gamma^2 n}.$$

The next corollary is now an immediate consequence of Proposition 5.2.3 combined with the previous Proposition 5.2.12 and we omit its proof.

Corollary 5.2.14 (Privatized radius of testing with hypercube sampling). Let $\alpha \in (0,1)$, $\gamma \in (0,1)$. Consider the family of tests $\left\{\Delta^{\mathrm{priv}}_{k,\alpha/2}\right\}, \alpha \in (0,1)$ defined in (5.2.5) and consider the hypercube mechanism $\mathbb{Q}_\gamma$. Let $\overline{A}_\alpha$ as in Proposition 5.2.3. Then, for all $A \geq \overline{A}_\alpha$ and all $k \in \mathbb{N}$ we obtain

$$\mathcal{R}\left(\Delta^{\mathrm{priv}}_{k,\alpha/2}, \mathbb{Q}_\gamma \mid \mathcal{E}^{\mathrm{R}}_{a_\bullet}, A\left(\rho_k \vee \rho_k^{\mathrm{HS}}\right)\right) \leq \alpha.$$

The previous corollary shows that compared with the (non-private) radius of testing ρ_k^2 derived in Section 3.3 the privatized upper bound has the additional term $(\rho_k^{\mathrm{HS}})^2$, where the variance term is increased.

Illustration 5.2.15 (Hypercube sampling). The upper bound for the radius of testing of the tests $\Delta^{\text{priv}}_{k,\alpha/2}, \alpha \in (0,1)$ and the hypercube sampling mechanism derived in Corollary 5.2.14 depend on the dimension parameter k. Analogously to Illustration 5.2.8 we define

$$\kappa_\star^{\text{HS}} := \arg\min_{k\in\mathbb{N}} \left\{ a_k^2 \vee c^2 \frac{k\nu_k^2}{\gamma^2 n} \right\} \tag{5.2.12}$$

and $\kappa_\star = \arg\min_{k\in\mathbb{N}} \left\{ a_k^2 \vee \frac{\nu_k^2}{n} \right\}$ as in (3.2.5), we can optimize the upper bound with respect to k and obtain the upper bound

$$\rho_\star^2 \vee \left(\rho_\star^{\text{HS}}\right)^2 \qquad \text{with} \qquad \rho_\star^{\text{HS}} := \min_{k\in\mathbb{N}} \rho_k^{\text{HS}}. \tag{5.2.13}$$

We illustrate the order of both terms under the typical smoothness and ill-posedness assumptions introduced in Illustration 3.2.6.

Order of the upper bound for the radius of testing under hypercube sampling			
a_j (smoothness)	$\lvert\varphi_j\rvert$ (ill-posedness)	$\rho_\star^2$	$(\rho_\star^{\text{HS}})^2$
j^{-s}	$\lvert j\rvert^{-p}$	$n^{-\frac{4s}{4s+4p+1}}$	$(\gamma^2 n)^{-\frac{4s}{4s+4p+3}}$
j^{-s}	$e^{-\lvert j\rvert^p}$	$(\log n)^{-\frac{2s}{p}}$	$(\log(\gamma^2 n))^{-\frac{2s}{p}}$
e^{-j^s}	$\lvert j\rvert^{-p}$	$n^{-1}(\log n)^{\frac{4p+1}{2s}}$	$(\gamma^2 n)^{-1}(\log(\gamma^2 n))^{\frac{4p+3}{2s}}$

Calculations for the risk bounds in Illustration 5.2.15. The order of $\rho_\star^2$ has already been established in Illustration 3.2.6. Consider $(\rho_\star^{\text{HS}})^2$.

1. **(ordinary smooth - mildly ill-posed)** The variance term $\frac{k\nu_k^2}{\gamma^2 n}$ is of order $\frac{k^{2p+3/2}}{\gamma^2 n}$ and the bias term a_k^2 is of order k^{-2s}. Hence, the optimal $\kappa_\star^{\text{HS}}$ satisfies $\kappa_\star^{\text{HS}} \sim (\gamma^2 n)^{\frac{2}{4s+4p+3}}$, which yields an upper bound of order $(\kappa_\star^{\text{HS}})^{-2s} \sim (\gamma^2 n)^{\frac{4s}{4s+4p+3}}$.

2. **(ordinary smooth - severely ill-posed)** The variance term $\frac{k\nu_k^2}{\gamma^2 n}$ is of order $\frac{k\exp(2k^p)}{\gamma^2 n}$. Hence, the optimal $\kappa_\star^{\text{HS}}$ satisfies $\kappa_\star^{\text{HS}} \sim (\log(\gamma^2 n/b_{\gamma^2 n}))^{1/p}$ with $b_n \sim (\log(\gamma^2 n))^{\frac{2s+1}{p}}$, which yields an upper bound of order $(\kappa_\star^{\text{HS}})^{-2s} \sim (\log(\gamma^2 n))^{-\frac{2s}{p}}$.

3. **(super smooth - mildly ill-posed)** The variance term $\frac{k\nu_k^2}{\gamma^2 n}$ is of order $\frac{k^{2p+3/2}}{\gamma^2 n}$ and the bias term a_k^2 is of order $\exp(-2k^s)$. Hence, the optimal $\kappa_\star^{\text{HS}}$ satisfies $\kappa_\star^{\text{HS}} \sim (\log(\gamma^2 n/b_{\gamma^2 n}))^{1/s}$ with $b_n \sim (\log(\gamma^2 n))^{\frac{4p+3}{2s}}$, which yields an upper bound of order $(\gamma^2 n)^{-1}(\log(\gamma^2 n))^{\frac{4p+3}{2s}}$.

□

Comparing our upper bounds in Illustration 5.2.15 with the known results in direct models (e.g. Lam-Weil et al. [2020]), we conjecture our radii to be optimal. In the next section (Perspectives) we take a first step to answer the important question: Is the deterioration of the radii of testing caused only by a poor choice of the privacy mechanism? Or is the attained deterioration unavoidable if we want to protect privacy? We provide an approach in form of a classical reduction scheme that might lead to a lower bound.

Perspectives

Lower bounds for testing under privacy constraints

A main difficulty when trying to prove lower bounds under privacy constraints is to characterize the required α-differential privacy property of the privacy mechanism in a way such that it can be exploited in the lower bound. A way to make the privacy constraint tangible is to write the privacy channels as operators, for which we can apply a singular value decomposition. The privacy constraint can be translated into a constraint on the singular values of this (appropriately defined) operator. Similar approaches in direct models have been considered in Lam-Weil et al. [2020] (for a direct density model) and Berrett and Butucea [2020] (for a direct model with discrete distributions). Let us give a preliminary framework for deriving a matching lower bound to the upper bound (5.2.13). Recall that the upper bound consists of the maximum of two terms $\rho_\star^2 \vee \left(\rho_\star^{\mathrm{LP}}\right)^2$ and we aim to prove separate lower bounds for these two situations. A matching lower bound for the first term has already been derived in Section 3.4. For the second term, we briefly outline the steps of our suggested lower bound framework.

1. **Reduction step.** For a fixed privacy channel $\mathbb{Q}$ standard reduction arguments (compare e.g. the proof of Proposition 3.4.1) show that

   $$\inf_{\Delta} \mathcal{R}\left(\Delta, \mathbb{Q} \mid \mathcal{E}, \rho\right) \geq 1 - \sqrt{\frac{\chi^2(\mathbb{P}_{\mu,\mathbb{Q}}, \mathbb{P}_{f^\circ,\mathbb{Q}})}{2}}$$

   where μ is a probability measure on the ρ-separated alternative and $\mathbb{P}_{\mu,\mathbb{Q}} = \int \mathbb{P}_{f,\mathbb{Q}} \mathrm{d}\mu(f)$.

2. **χ^2-divergence between privatized measures.** Straightforward calculations show that

   $$\chi^2(\mathbb{P}_{\mu,\mathbb{Q}}, \mathbb{P}_{f^\circ,\mathbb{Q}}) \leq \mathbb{E}_{\xi,\xi'} \exp\left(n \langle \boldsymbol{\Omega}_{\mathbb{Q}} g_\xi, g_{\xi'} \rangle_{\mathscr{L}^2}\right) - 1,$$

   where $\xi, \xi' \overset{\text{iid}}{\sim} \mu$ and we denote $g_\xi = \xi \circledast \varphi$, i.e. the density of Y in the case $X \sim \xi$. Moreover, we define the operator

   $$\begin{aligned} \boldsymbol{\Omega}_{\mathbb{Q}} &: \mathscr{L}^2 \longrightarrow \mathscr{L}^2 \\ &g \longmapsto \boldsymbol{\Omega}_{\mathbb{Q}}(g), \qquad \text{where} \qquad \boldsymbol{\Omega}_{\mathbb{Q}}(g)(\tilde{y}) := \int \Omega_{\mathbb{Q}}(y, \tilde{y}) g(y) \mathrm{d}y \end{aligned}$$

   with kernel $\Omega_{\mathbb{Q}} := \mathbb{E}_{f_0,\mathbb{Q}}\left(\frac{(\mathrm{q}(Z_j|y) - p_\circ(Z_j))(\mathrm{q}(Z_j|\tilde{y}) - p_\circ(Z_j))}{p_\circ^2(Z_j)}\right)$, where q is the density of $\mathbb{Q}$ w.r.t. some reference measure and $p_\circ = \int \mathrm{q}(\cdot \mid y) g^\circ(y) \mathrm{d}y$.

3. **Construction of the mixture measure μ.** Denote by $(\lambda_j)_{j\in\mathbb{N}}$ and $(v_j)_{j\in\mathbb{N}}$ the singular values and singular vectors of the operator $\boldsymbol{\Omega}_{\mathbb{Q}}$. Let $\eta \in \{\pm\}^k$ and $\theta \in \mathbb{R}^k$. Define

   $$\delta_\eta(x) := \sum_{l=1}^{k} \eta_l \theta_l v_l(x) \qquad \text{and} \qquad c_\eta = \int \delta_\eta(x) \mathrm{d}x$$

and the candidate functions

$$\xi_\eta(x) := g^\circ(x)(1 - c_\eta) + \delta_\eta(x).$$

The candidate functions ξ_η integrate to 1 by construction. It remains to verify that $\xi_\eta \geq 0$ (at least with high probability if η is sampled uniformly from $\{\pm\}^k$). Let μ be the uniform mixture on $\left\{\xi_\eta : \eta \in \{\pm\}^k\right\}$. If $\xi_\eta, \eta \in \{\pm\}^k$ are densities, then,

$$\chi^2\left(\mathbb{P}_{\mu,Q}, \mathbb{P}_{f_0,Q}\right) \leq \exp\left(\frac{n^2}{2}\sum_{m=1}^{k} \theta_m^4 \lambda_m^2\right) - 1. \tag{5.2.14}$$

4. **Control of the eigenvalues of $\boldsymbol{\Omega}_{\mathbb{Q}}$.** The γ-differentiable privacy constraint implies that any singular value λ_j of $\boldsymbol{\Omega}_{\mathbb{Q}}$ satisfies

$$\lambda_j \leq (e^\gamma - 1)^2.$$

The suggested steps are simply a rough outline, with several gaps still to be filled in. Arriving at (5.2.14) , the remaining demanding challenge is to construct $\theta \in \mathbb{R}^k$ such that μ is supported on the alternative and the χ^2-divergence in (5.2.14) is smaller than $4\alpha^2$.

Adaptive testing under local privacy constraints

Adaptivity in a local privacy setting seems to be a particularly interesting problem. As usual in nonparametric statistics the optimal dimension (in (5.2.12)) of our projection-type tests (considered in Chapter 5) relies on knowledge of smoothness properties of the unknown underlying density. Thus, there is a necessity to come up with adaptation procedures that do not rely on this knowledge, since it is generally not available in practise. Developing adaptive procedures in a local privacy setting is especially challenging, since the required steps need to be carried out by each data holder separately. Since statistical inference under privacy constraints is a relatively new field, there are so far only few results on adaptive strategies in general. For non-parametric estimation in a direct density model, Kroll [2019b] for instance investigates a privatized version of Lepski's method for kernel density bandwidth selection, where each data holder is asked to release evaluations of the kernel (scaled with a bandwidth) for the collection of bandwidths that appears in Lepski's method. The privatization mechanisms have to be scaled appropriately such that privacy protection is still guaranteed although more data (i.e. evaluations of a kernel for several bandwidths) is released. Let us briefly explain what makes adaptivity challenging in the context of nonparametric testing. In order to apply the classical Bonferroni aggregation method to the privatized tests, the statistician needs access to the entire collection of tests $\Delta_k, k \in \mathcal{K}$, where k is the tuning parameter and $\mathcal{K}$ a collection of such parameters, over which one wishes to aggregate. We have seen, however, that our proposed privatization strategies highly depend on the dimension parameter k. To be able to describe the problem at hand more accurately, let us briefly recall the main idea. In both hypercube sampling and Laplace perturbation, the j-th data holder privatizes the observation Y_j by evaluating the first k basis functions $\{e_m(Y_j)\}_{m \in [\![k]\!]}$ and transforming these k-dimensional vectors by the chosen privatization mechanism. Since the same observation Y_j is used in the evaluation of each basis function, the privatization mechanism naturally depends on the number of evaluated basis functions, which we want to release. This is a common phenomenon, e.g. also noted in Lam-Weil et al. [2020]. The influence of the dimension k on the privatized samples is reflected in their variance after privatization. For Laplace perturbation the variance of the (γ-)privatized observations is of order $\sigma_{\mathrm{LP}} \sim \frac{k}{\gamma}$, whereas the dependence on k is reduced to $\sigma_{\mathrm{HS}} \sim \frac{\sqrt{k}}{\gamma}$ for hypercube sampling. Our general upper bounds for the radius of testing demonstrate the effect that the dependence of the variance on k has on the radii. In fact, the upper bounds (with the notation of Chapter 5) for a projection-based test with dimension k and a privatization mechanism yielding private views with variance σ (typically depending on k) are given by

$$(\rho_{k,\sigma}^{\mathrm{priv}})^2 := a_k^2 \vee (1+\sigma^2)\frac{\nu_k^2}{n},$$

which have to be minimised with respect to k in order to be optimal. Without knowledge on the true underlying smoothness structure, i.e. on $a_\bullet$, we can a priori only derive rough upper bounds for the optimal dimension. These upper bounds are – in terms of the sample size n – typically of order n^c for some $c > 0$ (in mildly ill-posed models) or of order $\log n$ (in severely ill-posed models). Hence, the naive approach is to aggregate over an (appropriate) class containing all possibly optimal dimension parameters (i.e. up to the upper bound). To do so, the data holder would have to release privatized version of the evaluated basis functions up to the upper bound for the optimal dimension. Therefore, additionally to the usual cost to pay for the protection of privacy and the deterioration due to the aggregation, there is a factor caused by the necessary privatisation for all dimensions in the collection $\mathcal{K}$. It is an open question whether this strategy then still exhibits optimal adaptive behaviour or whether completely different adaptation strategies and privatization mechanisms need to be developed.

Bibliography

P. Alquier, E. Gautier, and G. Stoltz. *Inverse problems and high-dimensional estimation: Stats in the Château Summer School, August 31-September 4, 2009*, volume 203. Springer Science & Business Media, 2011.

S. Balakrishnan and L. Wasserman. Hypothesis testing for high-dimensional multinomials: A selective review. *The Annals of Applied Statistics*, 12(2):727–749, 2018.

S. Balakrishnan and L. Wasserman. Hypothesis testing for densities and high-dimensional multinomials: Sharp local minimax rates. *The Annals of Statistics*, 47(4):1893–1927, 2019.

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

R. F. Barber and J. C. Duchi. Privacy and statistical risk: Formalisms and minimax bounds. *arXiv preprint arXiv:1412.4451*, 2014.

J. Baumeister and A. Leitao. *Topics in inverse problems*. IMPA, 2005.

T. B. Berrett and C. Butucea. Locally private non-asymptotic testing of discrete distributions is faster using interactive mechanisms. *arXiv preprint arXiv:2005.12601*, 2020.

P. J. Bickel and Y. Ritov. Estimating integrated squared density derivatives: sharp best order of convergence estimates. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 381–393, 1988.

L. Birgé. An alternative point of view on Lepski's method. *Institute of Mathematical Statistics Lecture Notes - Monograph Series*, 36:113–133, 2001. ISSN 0749-2170.

L. Birgé, Lucien and P. Massart. Estimation of integral functionals of a density. *The Annals of Statistics*, 23(1):11–29, 02 1995.

C. Butucea. Goodness-of-fit testing and quadratic functional estimation from indirect observations. *The Annals of Statistics*, 35(5):1907–1930, 2007.

C. Butucea and F. Comte. Adaptive estimation of linear functionals in the convolution model and applications. *Bernoulli*, 15(1):69–98, 2009.

C. Butucea and K. Meziani. Quadratic functional estimation in inverse problems. *Statistical Methodology*, 8(1):31–41, 2011.

C. Butucea, C. Matias, and C. Pouet. Adaptivity in convolution models with partially known noise distribution. *Electronic Journal of Statistics*, 2:897–915, 2008.

C. Butucea, C. Matias, and C. Pouet. Adaptive goodness-of-fit testing from indirect observations. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 45:352–372, 2009.

C. Butucea, A. Dubois, M. Kroll, and A. Saumard. Local differential privacy: Elbow effect in optimal density estimation and adaptation over Besov ellipsoids. *Bernoulli*, 26(3):1727–1764, 2020.

T. T. Cai and M. G. Low. A note on nonparametric estimation of linear functionals. *The Annals of Statistics*, 31(4):1140–1153, 2003.

T. T. Cai and M. G. Low. Minimax estimation of linear functionals over nonconvex parameter spaces. *The Annals of Statistics*, 32(2):552–576, 2004.

T. T. Cai and M. G. Low. Nonquadratic estimators of a quadratic functional. *The Annals of Statistics*, 33(6):2930–2956, 2005a.

T. T. Cai and M. G. Low. Adaptive estimation of linear functionals under different performance measures. *Bernoulli*, 11(2):341–358, 2005b.

T. T. Cai and M. G. Low. On adaptive estimation of linear functionals. *The Annals of Statistics*, 33(5):2311–2343, 2005c.

T. T. Cai and M. G. Low. Optimal adaptive estimation of a quadratic functional. *The Annals of Statistics*, 34(5):2298–2325, 2006.

A. Carpentier and N. Verzelen. Adaptive estimation of the sparsity in the Gaussian vector model. *The Annals of Statistics*, 47(1):93–126, 2019.

M. Carrasco, J.-P. Florens, and E. Renault. Linear inverse problems in structural econometrics estimation based on spectral decomposition and regularization. *Handbook of econometrics*, 6: 5633–5751, 2007.

L. Cavalier. Nonparametric statistical inverse problems. *Inverse Problems*, 24(3):034004, 2008.

L. Cavalier and N. W. Hengartner. Adaptive estimation for inverse problems with noisy operators. *Inverse problems*, 21(4):1345, 2005.

L. Cavalier and A. Tsybakov. Sharp adaptation for inverse problems with random noise. *Probability Theory and Related Fields*, 123(3):323–354, 2002.

L. Cavalier, G. Golubev, D. Picard, and A. Tsybakov. Oracle inequalities for inverse problems. *The Annals of Statistics*, 30(3):843–874, 2002.

L. Cavalier, Y. Golubev, O. Lepski, and A. Tsybakov. Block thresholding and sharp adaptive estimation in severely ill-posed inverse problems. *Theory of Probability & Its Applications*, 48 (3):426–446, 2004.

C. Chesneau. On adaptive wavelet estimation of a quadratic functional from a deconvolution problem. *Annals of the Institute of Statistical Mathematics*, 63(2):405–429, 2011.

O. Collier, L. Comminges, and A. B. Tsybakov. Minimax estimation of linear and quadratic functionals on sparsity classes. *The Annals of Statistics*, 45(3):923–958, 2017.

O. Collier, L. Comminges, A. B. Tsybakov, and N. Verzelen. Optimal adaptive estimation of linear functionals under sparsity. *The Annals of Statistics*, 46(6A):3130–3150, 2018.

F. Comte and C. Lacour. Data-driven density estimation in the presence of additive noise with unknown distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):601–627, 2011.

F. Comte and M.-L. Taupin. Adaptive density deconvolution for circular data. http://helios.mi.parisdescartes.fr/ comte/efrodens5.pdf, 2003.

D. Cox and P. Lewis. The statistical analysis of series of events. *Methuen & Co ltd, London*, 1966.

A. DasGupta. *Asymptotic theory of statistics and probability.* Springer Science & Business Media, 2008.

I. Dinur and K. Nissim. Revealing information while preserving privacy. In *Proceedings of the twenty-second ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pages 202–210, 2003.

J. C. Duchi, M. I. Jordan, and M. J. Wainwright. Minimax optimal procedures for locally private estimation. *Journal of the American Statistical Association*, 113(521):182–201, 2018.

C. Dwork. Differential privacy, in automata, languages and programming. *Lecture Notes in Computer Sciencce*, 4052:112, 2006.

C. Dwork. Differential privacy: A survey of results. In *International conference on theory and applications of models of computation*, pages 1–19. Springer, 2008.

C. Dwork and K. Nissim. Privacy-preserving datamining on vertically partitioned databases. In *Annual International Cryptology Conference*, pages 528–544. Springer, 2004.

S. Efromovich. Density estimation for the case of supersmooth measurement error. *Journal of the American Statistical Association*, 92(438):526–535, 1997.

S. Efromovich and V. Koltchinskii. On inverse problems with unknown operators. *IEEE Transactions on Information Theory*, 47(7):2876–2894, 2001.

H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems*, volume 375. Springer Science & Business Media, 1996.

M. Ermakov. On optimal solutions of the deconvolution problem. *Inverse problems*, 6(5):863, 1990.

M. Ermakov. Minimax detection of a signal in the heteroscedastic Gaussian white noise. *Journal of Mathematical Sciences*, 137(1):4516–4524, 2006.

J. Fan. On the optimal rates of convergence for nonparametric deconvolution problems. *The Annals of Statistics*, pages 1257–1272, 1991.

N. I. Fisher. *Statistical analysis of circular data.* Cambridge University Press, 1995.

M. Fromont and B. Laurent. Adaptive goodness-of-fit tests in a density model. *The Annals of Statistics*, 34(2):680–720, 2006.

J. Gill and D. Hangartner. Circular data in political science and how to handle it. *Political Analysis*, 18(3):316–336, 2010.

E. Gine and R. Nickl. *Mathematical Foundations of Infinite-Dimensional Statistical Models.* Cambridge University Press, 2015.

A. Goldenshluger and S. V. Pereverzev. Adaptive estimation of linear functionals in Hilbert scales from indirect white noise observations. *Probability Theory and Related Fields*, 118(2): 169–186, 2000.

A. Goldenshluger and S. V. Pereverzev. On adaptive inverse estimation of linear functionals in Hilbert scales. *Bernoulli*, 9(5):783–807, 2003.

G. Golubev. On adaptive estimation of linear functionals from observations against white noise. *Problems of Information Transmission*, 56(2):185–200, 2020.

J. Hadamard. Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton university bulletin*, pages 49–52, 1902.

R. Hall, A. Rinaldo, and L. Wasserman. Differential privacy for functions and functional data. *Journal of Machine Learning Research*, 14(Feb):703–727, 2013.

M. Hoffmann and M. Reiss. Nonlinear estimation for linear inverse problems with error in the operator. *The Annals of Statistics*, 36(1):310–336, 2008.

I. Ibragimov and R. Khas' minskii. Estimating the value of a linear functional of a distribution density. *Journal of Soviet Mathematics*, 44(4):454–465, 1989.

I. A. Ibragimov and R. Z. Khas' minskii. On nonparametric estimation of the value of a linear functional in Gaussian white noise. *Theory of Probability & Its Applications*, 29(1):18–32, 1985.

I. A. Ibragimov and R. Z. Khas' minskii. Estimation of linear functionals in Gaussian noise. *Theory of Probability & Its Applications*, 32(1):30–39, 1988.

Y. Ingster. Asymptotically minimax hypothesis testing for nonparametric alternatives I. *Mathematical Methods of Statistics*, 2(2):85–114, 1993a.

Y. Ingster. Asymptotically minimax hypothesis testing for nonparametric alternatives II. *Mathematical Methods of Statistics*, 2(2):171–189, 1993b.

Y. Ingster. Asymptotically minimax hypothesis testing for nonparametric alternatives III. *Mathematical Methods of Statistics*, 2(2):249—268, 1993c.

Y. Ingster and I. A. Suslina. *Nonparametric goodness-of-fit testing under Gaussian models*, volume 169. Springer Science & Business Media, 2012.

Y. I. Ingster, T. Sapatinas, and I. A. Suslina. Minimax signal detection in ill-posed inverse problems. *The Annals of Statistics*, 40(3):1524–1549, 2012a.

Y. I. Ingster, T. Sapatinas, and I. A. Suslina. Supplement to the article: Minimax signal detection in ill-posed inverse problems. *The Annals of Statistics*, 40(3):1524–1549, 2012b.

J. Johannes and M. Schwarz. Adaptive Gaussian inverse regression with partially unknown operator. *Communications in Statistics-Theory and Methods*, 42(7):1343–1362, 2013.

I. M. Johnstone and B. W. Silverman. Speed of estimation in positron emission tomography and related inverse problems. *The Annals of Statistics*, pages 251–280, 1990.

G. Kerkyacharian, T. M. P. Ngoc, and D. Picard. Localized spherical deconvolution. *The Annals of Statistics*, 39(2):1042–1068, 2011.

M. Kroll. Rate optimal estimation of quadratic functionals in inverse problems with partially unknown operator and application to testing problems. *ESAIM: Probability and Statistics*, 23:524–551, 2019a.

M. Kroll. Pointwise adaptive kernel density estimation under local approximate differential privacy. *arXiv preprint arXiv:1907.06233*, 2019b.

C. Lacour and T. M. P. Ngoc. Goodness-of-fit test for noisy directional data. *Bernoulli*, 20(4): 2131–2168, 2014.

J. Lam-Weil, B. Laurent, and J.-M. Loubes. Minimax optimal goodness-of-fit testing for densities under a local differential privacy constraint. *arXiv preprint arXiv:2002.04254*, 2020.

B. Laurent. Adaptive estimation of a quadratic functional of a density by model selection. *ESAIM: Probability and Statistics*, 9:1–18, 2005.

B. Laurent and P. Massart. Adaptive estimation of a quadratic functional by model selection. *The Annals of Statistics*, 28(5):1302–1338, 2000.

B. Laurent, S. Huet, and Y. Baraud. Adaptive tests of linear hypotheses by model selection. *The Annals of Statistics*, 31(1):225–251, 2003.

B. Laurent, C. Ludena, and C. Prieur. Adaptive estimation of linear functionals by model selection. *Electronic journal of statistics*, 2:993–1020, 2008.

B. Laurent, J.-M. Loubes, and C. Marteau. Testing inverse problems: a direct or an indirect problem? *Journal of Statistical Planning and Inference*, 141(5):1849–1861, 2011.

B. Laurent, J.-M. Loubes, and C. Marteau. Non asymptotic minimax rates of testing in signal detection with heterogeneous variances. *Electronic Journal of Statistics*, 6:91–122, 2012.

O. V. Lepski and V. G. Spokoiny. Minimax nonparametric hypothesis testing: the case of an inhomogeneous alternative. *Bernoulli*, 5(2):333–358, 1999.

J.-M. Loubes and C. Marteau. Goodness-of-fit testing strategies from indirect observations. *Journal of Nonparametric Statistics*, 26(1):85–99, 2014.

B. A. Mair and F. H. Ruymgaart. Statistical inverse estimation in Hilbert scales. *SIAM Journal on Applied Mathematics*, 56(5):1424–1444, 1996.

K. V. Mardia. *Statistics of directional data.* Academic press, 1972.

K. V. Mardia and P. E. Jupp. *Directional statistics*, volume 494. John Wiley & Sons, 2009.

C. Marteau. Regularization of inverse problems with unknown operator. *Mathematical Methods of Statistics*, 15(4):415–443, 2006.

C. Marteau and P. Mathé. General regularization schemes for signal detection in inverse problems. *Mathematical Methods of Statistics*, 23(3):176–200, 2014.

C. Marteau and T. Sapatinas. A unified treatment for non-asymptotic and asymptotic approaches to minimax signal detection. *Statistics Surveys*, 9:253–297, 2015.

C. Marteau and T. Sapatinas. Minimax goodness-of-fit testing in ill-posed inverse problems with partially unknown operators. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 53(4):1675–1718, 2017a.

C. Marteau and T. Sapatinas. Minimax signal detection under weak noise assumptions. *Mathematical Methods of Statistics*, 26(4):282–298, 2017b.

P. Mathé and S. V. Pereverzev. Optimal discretization of inverse problems in Hilbert scales. regularization and self-regularization of projection methods. *SIAM Journal on Numerical Analysis*, 38(6):1999–2021, 2001.

P. Mathé and S. V. Pereverzev. Direct estimation of linear functionals from indirect noisy observations. *Journal of Complexity*, 18(2):500–516, 2002.

F. McSherry and K. Talwar. Mechanism design via differential privacy. In *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07)*, pages 94–103. IEEE, 2007.

A. Meister. Deconvolution problems in nonparametric statistics. *Lecture Notes in Statistics*, 2009.

M. H. Neumann. On the effect of estimating the error density in nonparametric deconvolution. *Journal of Nonparametric Statistics*, 7(4):307–330, 1997.

S. Schluttenhofer and J. Johannes. Adaptive minimax testing in inverse Gaussian sequence space models. *arXiv preprint arXiv:2002.07623*, 2020a.

S. Schluttenhofer and J. Johannes. Minimax testing and quadratic functional estimation for circular convolution. *arXiv preprint arXiv:2004.12714*, 2020b.

S. Schluttenhofer and J. Johannes. Adaptive minimax testing for circular convolution. *arXiv preprint arXiv:2007.06388*, 2020c.

R. J. Serfling. *Approximation theorems of mathematical statistics*, volume 162. John Wiley & Sons, 2009.

V. Spokoiny. Adaptive hypothesis testing using wavelets. *The Annals of Statistics*, 24(6):2477–2498, 1996.

L. A. Stefanski and R. J. Carroll. Deconvolving kernel density estimators. *Statistics*, 21(2): 169–184, 1990.

A. B. Tsybakov. Optimal aggregation of classifiers in statistical learning. *The Annals of Statistics*, 32(1):135–166, 2004.

A. B. Tsybakov. *Introduction to Nonparametric Estimation.* Springer New York, 2009.

L. Wasserman and S. Zhou. A statistical framework for differential privacy. *Journal of the American Statistical Association*, 105(489):375–389, 2010.

Y. Wei and M. J. Wainwright. The local geometry of testing in ellipses: Tight control via localized Kolmogorov widths. *IEEE Transactions on Information Theory*, 2020.

D. Werner. *Funktionalanalysis.* Springer, 2006.