

RATE OPTIMAL ESTIMATION OF QUADRATIC FUNCTIONALS IN INVERSE PROBLEMS WITH PARTIALLY UNKNOWN OPERATOR AND APPLICATION TO TESTING PROBLEMS^{*,**}

MARTIN KROLL^{***}

Abstract. We consider the estimation of quadratic functionals in a Gaussian sequence model where the eigenvalues are supposed to be unknown and accessible through noisy observations only. Imposing smoothness assumptions both on the signal and the sequence of eigenvalues, we develop a minimax theory for this problem. We propose a truncated series estimator and show that it attains the optimal rate of convergence if the truncation parameter is chosen appropriately. Consequences for testing problems in inverse problems are equally discussed: in particular, the minimax rates of testing for signal detection and goodness-of-fit testing are derived.

Mathematics Subject Classification. 62G05, 62G10.

Received July 4, 2018. Accepted December 4, 2018.

1. INTRODUCTION

1.1. Problem statement

We consider the Gaussian sequence model

$$X_j = \lambda_j \theta_j + \varepsilon \xi_j, \quad j \in \mathbb{N}^* = \{1, 2, \dots\}, \quad \text{and} \quad (1.1)$$

$$Y_j = \lambda_j + \sigma \eta_j, \quad j \in \mathbb{N}^* \quad (1.2)$$

where $(\xi_j)_{j \in \mathbb{N}^*}$, $(\eta_j)_{j \in \mathbb{N}^*}$ are independent random vectors with independent standard Gaussian components and $\varepsilon, \sigma \in (0, 1)$ are known noise levels. Given some known and fixed ‘reference point’ $(\theta_j^o)_{j \in \mathbb{N}^*}$, we will in this work

* This work has been carried out within the framework of the ECODEC laboratory of excellence, bearing the reference ANR-11-LABX-0047.

** The author is grateful to two anonymous referees for their detailed comments that led to an essential improvement of the article.

Keywords and phrases: Inverse problem, unknown eigenvalues, minimax theory, rate optimality, truncated series estimator, non-parametric testing, goodness-of-fit testing, signal detection.

ENSAE-ParisTech CREST, 5 avenue Henry Le Chatelier, 91120 Palaiseau, France.

*** Corresponding author: martin.kroll@ensae.fr

address the following two questions:

1. Let $(\omega_j)_{j \in \mathbb{N}^*}$ be some known sequence of weights. How can we estimate the value of the quadratic functional

$$q(\theta) = \sum_{j=1}^{\infty} \omega_j^2 (\theta_j - \theta_j^\circ)^2$$

from data $(X_j)_{j \in \mathbb{N}^*}$ and $(Y_j)_{j \in \mathbb{N}^*}$ in an optimal way?

2. How can we test the null hypothesis $\theta = \theta^\circ$ against the alternative $\theta \in \Theta_1$ for some Θ_1 with $\theta^\circ \notin \Theta_1$?

Concerning both questions, the sequence $(\lambda_j)_{j \in \mathbb{N}^*}$ is a nuisance parameter and only accessible by means of the observations $(Y_j)_{j \in \mathbb{N}^*}$. Specific choices include both the case $\lambda_j \equiv 1$ (then, (1.1) is the classical Gaussian sequence model with *direct* observations) and the case $\lambda_j \rightarrow 0$ making the inverse problem of reconstructing θ *ill-posed* (see [8], Def. 1.1 for a definition of well-/ill-posedness). Precise assumptions on all model parameters will be given in Section 2. To the best of our knowledge, the model given by (1.1) and (1.2) was introduced explicitly in [9] for the first time, and is also referred to as an inverse problem with *partially unknown operator* [26, 35]. In the context of inverse problems, in its general form given by an operator equation $X = A\theta + \varepsilon\xi$, this model provides something between the classical assumption that the operator A is known [8, 13] and the assumption that the operator is only accessible by a blurred observation $Y = A + \sigma\Xi$ [15, 21]: it arises by the structural assumption that the operator A^*A is diagonal with eigenvalues $(\lambda_j^2)_{j \in \mathbb{N}^*}$. We refer to the references [9, 26, 35] for a more detailed derivation and further motivation of the model. Note that, whereas the non-parametric estimation of the parameter θ itself from observations (1.1) and (1.2) (including adaptation) was intensively studied in [9, 26], the estimation of quadratic functionals has not yet been considered, and also the question of non-parametric testing has been investigated only recently (see the following Subsection 1.2 for a discussion of related work).

1.2. Related work

Starting with the paper [1], the estimation of quadratic functionals has received a lot of attention in non-parametric statistics, in particular in models with direct observations [6, 7, 11, 14, 17, 19, 20, 27–29, 32, 33, 37]. In the context of statistical inverse problems, there is much less work dedicated to this problem. Butucea [3] provides a goodness-of-fit test in a convolution model where the test statistic is based on the estimator of a quadratic functional. The paper [5] considers observations as in (1.1) but assumes the sequence of eigenvalues to be known. Under this assumption, minimax upper bounds in terms of ε are derived for both ordinary smooth and supersmooth θ . In addition, the authors assume that their approach even provides optimal constants. Chesneau [10] considers adaptive estimation of the \mathbb{L}^2 -norm in a model where a convolution product of an unknown function and a known function is corrupted by Gaussian noise.

The estimation of quadratic functionals is closely related to hypotheses testing since estimators of quadratic functionals provide natural building blocks for test statistics. Starting with the seminal paper [16], the theory of non-parametric testing in direct Gaussian sequence space models has been rigorously developed in a series of papers by Ingster [23] (see also the monograph [22]). In the domain of inverse problems, an increasing interest in theoretical results in the spirit of the book [22] has arisen within the last decade [24, 31, 34, 35], partially motivated by applications coming from biology [2] or astrophysics [30]. However, concerning inverse problems with partially unknown operator, the existing research literature reduces to the paper [35] that considers the same model as in the present work. In contrast to our approach in Section 5 where we use the sum of type I and type II error in order to measure the performance of tests, the authors of [35] consider level- α -tests (*i.e.*, tests whose type I error is bounded from above by some prespecified $\alpha \in (0, 1)$) and try to minimize the type II error under this constraint. In this framework, the authors derive upper and lower bounds for the so-called *separation rate*. Their test statistic is also based on estimation of a quadratic functional but only the goodness-of-fit testing problem is considered. The authors obtain a slight gap by a logarithmic factor between upper and lower bounds with respect to the noise level σ . A main difference between the present paper and [35] concerns

the minimax methodology: Marteau and Sapatinas [35] impose a smoothness condition on the sequence $(\lambda_j)_{j \in \mathbb{N}^*}$ (equivalent to our one introduced in Sect. 2) only in order to establish lower bounds, but the construction of their test statistic is independent of this smoothness. Thus, their testing procedure is adaptive with respect to the sequence $(\lambda_j)_{j \in \mathbb{N}}$, whereas we assume the order of the decay of this sequence to be known. Imposing this additional assumption, we are able to derive upper and lower bounds for the testing rate that match (without any logarithmic gap). It might be of interest to explore to what extent the extra logarithmic factors in [35] might be unavoidable in the adaptive scenario. However, answering this question is outside the scope of this work and deferred to future research.

1.3. Organisation and main contributions of the paper

Let us summarize the main contributions of this paper. We emphasize in advance that all results of the paper are non-asymptotic.

- We introduce a truncated series estimator of $q(\theta)$ (Sect. 2), and derive minimax upper bounds for this estimator in terms of the noise levels ε and σ (Sect. 3). The construction of the estimator is based on a *sample cloning* technique that has not been used before to construct estimators of quadratic functionals.
- We prove minimax lower bounds for the estimation of $q(\theta)$ from data (1.1) and (1.2). These results show that the truncated series estimator is rate optimal provided that the truncation parameter is chosen appropriately.
- Our abstract results indicate an ‘elbow effect’ of the optimal rate of convergence in terms of the noise level σ that is similar to the well-known elbow effect in ε [17]. However, the rate in σ is in general faster than the one in ε and the parametric rate σ^2 can be attained in cases where the non-parametric regime holds with respect to ε . For instance, in the case that the signal belongs to a Sobolev class of index p and the considered inverse problem is mildly ill-posed with degree of ill-posedness equal to a , the optimal rate of convergence will turn out to be

$$\varepsilon^2 \vee \varepsilon^{16p/(4a+4p+1)} \vee \sigma^2 \vee \sigma^{4p/a}.$$

- In Section 5, as a rather direct application of our results on the estimation of quadratic functionals we consider non-parametric testing problems of the type

$$\mathcal{H}_0: \theta = \theta^\circ \in \Theta \quad \text{against} \quad \mathcal{H}_1: \theta \in \Theta, \|\theta - \theta^\circ\|_2 \geq r$$

for some $r > 0$. As already remarked by Marteau and Sapatinas [35], the case of signal detection ($\theta^\circ = 0$, Sect. 5.1) and the one of goodness-of-fit testing ($\theta^\circ \neq 0$, Sect. 5.2) have to be treated separately. For both problems, we derive the minimax rate of testing and propose a test statistic attaining this rate. In particular, in coincidence with the findings in [31], it turns out that for the signal detection problem it is sufficient to consider the observation (1.1) and construct a test statistic based on an estimator of a quadratic functional of $\tilde{\theta} = \lambda\theta$. For the goodness-of-fit problem, however, the testing rate depends also on σ and both observations, X and Y , are taken into account for the construction of the test statistic.

2. METHODOLOGY

2.1. Notation

We frequently denote entire sequences by single letters when writing ‘the sequence a ’ instead of ‘the sequence $(a_j)_{j \in \mathbb{N}^*}$ ’. Numerical operations on sequences like a^{-1} are to be understood elementwise. Throughout C denotes a purely numerical constant and $C(\dots)$ a constant that depends only on the parameters indicated within parentheses. $x \lesssim y$ is shorthand for $x \leq Cy$, and we write $x \asymp y$ if $x \lesssim y$ and $y \lesssim x$ hold simultaneously. Moreover, $x \asymp_\nu y$ means that $x\nu^{-1} \leq y \leq \nu x$. We put $\llbracket x, y \rrbracket = [x, y] \cap \mathbb{Z}$ for $x, y \in \mathbb{R}$.

2.2. Truncated series estimator

In order to define a truncated series estimators, we first generate two independent instances of the Y sample by the following *sample cloning* technique which is well-known in the context of aggregation (see [39], Lem. 2.1): let $\tilde{\eta}$ be a sequence of independent standard Gaussian random variables independent of ξ and η . For $j \in \mathbb{N}^*$, we put

$$\tilde{Y}_j = Y_j + \sigma\tilde{\eta}_j \quad \text{and} \quad \bar{Y}_j = Y_j - \sigma\tilde{\eta}_j.$$

Then \tilde{Y}_j, \bar{Y}_j are i.i.d. $\mathcal{N}(\lambda_j, 2\sigma^2)$, and the price to pay for the availability of two independent samples is a doubling of the variance. Based on the availability of the samples $\tilde{Y} = (\tilde{Y}_j)_{j \in \mathbb{N}^*}, \bar{Y} = (\bar{Y}_j)_{j \in \mathbb{N}^*}$ we define, for any $k \in \mathbb{N}^*$, the truncated series estimator

$$\hat{\mathbf{q}}_k = \sum_{j=1}^k \omega_j^2 \frac{U_j}{V_j} \mathbf{1}_{\Omega_j} \tag{2.1}$$

where $U_j := (X_j - \tilde{Y}_j\theta_j^\circ)^2 - \varepsilon^2 - 2(\theta_j^\circ)^2\sigma^2$, $V_j := \bar{Y}_j^2 - 2\sigma^2$ and $\Omega_j := \{\bar{Y}_j^2 \geq 3\sigma^2\}$. Note that U_j and V_j are unbiased estimators of $\lambda_j^2(\theta_j - \theta_j^\circ)^2$ and λ_j^2 , respectively, guaranteeing that the fraction U_j/V_j is at least a consistent estimator of $(\theta_j - \theta_j^\circ)^2$. In addition, due to the construction based on sample cloning, U_j and $\mathbf{1}_{\Omega_j}/V_j$ are independent as they can be written as $U_j = f(X_j, \tilde{Y}_j)$ and $\mathbf{1}_{\Omega_j}/V_j = g(\bar{Y}_j)$ for non-random functions f and g . Inspired by [36], the additional cut-off $\mathbf{1}_{\Omega_j}$ in (2.1) excludes too small values of V_j that would otherwise lead to an unstable behaviour of the entire estimator. As usual in non-parametric statistics, the value of the truncation parameter $k \in \mathbb{N}^*$ has to be chosen by the statistician and crucially effects the performance of the estimator. In Section 3, we first derive an upper risk bound for $\hat{\mathbf{q}}_k$ that holds for any $k \in \mathbb{N}^*$, and then take the minimizer of this bound to define our final estimator. This specific choice will turn out to define a rate optimal estimator under mild assumptions (of course, the resulting estimator is not adaptive). Let us note that, in order to derive a minimax optimal estimator only, other truncated series estimators could have been chosen. The construction of our estimator, however, is motivated by our application to testing in Section 5.

2.3. Minimax estimation

Given sequences γ and α , let us define the ℓ^2 -ellipsoid

$$\Theta = \Theta(\gamma, L) = \left\{ \theta \in \ell^2 : \sum_{j=1}^{\infty} \gamma_j^2 \theta_j^2 \leq L^2 \right\}$$

and the ℓ^2 -hyperrectangle

$$\mathcal{E} = \mathcal{E}(\alpha, d) = \{ \lambda \in \ell^\infty : d^{-1}\alpha_j \leq |\lambda_j| \leq d\alpha_j \}.$$

We usually suppress the dependence of Θ and \mathcal{E} on γ, α, L, d in the notation. For the rest of the paper, we assume that $(\theta, \lambda) \in \Theta \times \mathcal{E}$.

Definition 2.1 (Minimax rate of estimation, minimax estimator). An estimator $\hat{\mathbf{q}}$ of $\mathbf{q}(\theta)$ attains the rate $\psi_{\varepsilon, \sigma}^2$ over the smoothness classes Θ and \mathcal{E} if there exists a numerical constant $C > 0$ such that

$$\sup_{\theta \in \Theta} \sup_{\lambda \in \mathcal{E}} \mathbf{E}[(\hat{\mathbf{q}} - \mathbf{q}(\theta))^2] \leq C\psi_{\varepsilon, \sigma}^2.$$

The rate $\psi_{\varepsilon,\sigma}^2$ is called *minimax optimal* if in addition

$$\inf_{\tilde{q}} \sup_{\theta \in \Theta} \sup_{\lambda \in \mathcal{E}} \mathbf{E}[(\tilde{q} - q(\theta))^2] \geq c\psi_{\varepsilon,\sigma}^2 \quad (2.2)$$

holds for some $c > 0$ where the infimum is taken over all estimators based on observations (1.1) and (1.2).

In this work, the minimax optimal rate is derived under the following assumption on the sequences α, γ and ω .

Assumption 2.2. The sequences α and $\omega\gamma^{-1}$ are non-increasing and normalized such that $\alpha_1 = \gamma_1 = \omega_1 = 1$.

Assumption 2.2 is rather mild and satisfied by all the examples considered later. The proof of Theorem 3.1 shows that $\omega_k^4 \gamma_k^{-4}$ is the order of the squared bias of our estimator, and hence the convergence of $\omega\gamma^{-1}$ to zero ensures consistency as $\max\{\varepsilon, \sigma\}$ tends to zero. The following special choices of the sequences α and γ satisfy Assumption 2.2, and will be used throughout the paper to illustrate the general results. Concerning the sequence α we consider either

- the case $\alpha_j \asymp j^{-a}$ for some $a > 0$ (the inverse problem is *mildly ill-posed* and a the *degree of ill-posedness*), or
- the case $\alpha_j \asymp \exp(-ja)$ for some $a > 0$ (the inverse problem is *severely ill-posed*).

Concerning the sequence γ we consider either

- the case $\gamma_j \asymp j^p$ for some $p > 0$ (Θ is a *Sobolev ellipsoid*), or
- the case $\gamma_j \asymp \exp(pj)$ for some $p > 0$ (Θ is an *ellipsoid of analytic functions*).

The same smoothness assumptions have equally been used for the purpose of illustration in [24, 35] making our results directly comparable to the ones obtained in those papers.

2.4. Minimax theory of testing

In Section 5, we consider the problem of testing the simple hypothesis $\theta = \theta^\circ$ against the composite alternative $\theta \in \Theta_1$ with $\theta^\circ \notin \Theta_1$ (more precisely, we test $(\theta, \lambda) \in \{\theta^\circ\} \times \mathcal{E}$ against $(\theta, \lambda) \in \Theta_1 \times \mathcal{E}$). Usually, the case $\theta^\circ = 0$ is referred to as *signal detection* and the case $\theta^\circ \neq 0$ as *goodness-of-fit testing*. By definition, a *test statistic* Δ is a $\{0, 1\}$ -valued function based on the observations (X, Y) . Its performance is measured by the sum of type I and maximal type II error, $\mathbf{P}_0(\Delta = 1) + \sup_{\theta \in \Theta_1} \mathbf{P}_\theta(\Delta = 0)$, and the corresponding benchmark is the quantity

$$\inf_{\tilde{\Delta}} \left\{ \mathbf{P}_0(\tilde{\Delta} = 1) + \sup_{\theta \in \Theta_1} \mathbf{P}_\theta(\tilde{\Delta} = 0) \right\}$$

where the infimum is taken over all test statistics $\tilde{\Delta}$. It is well-known that, apart from smoothness assumptions, the null hypothesis θ° must be separated from the alternative Θ_1 at least by a certain distance in order to make non-trivial testing possible. In this spirit, we consider for $r > 0$ the testing problems

$$\mathcal{H}_0: \theta = \theta^\circ \in \Theta \quad \text{against} \quad \mathcal{H}_1: \theta - \theta^\circ \in \Theta_1(r)$$

where $\Theta_1(r) = \Theta \cap \{\theta \in \ell^2(\mathbb{N}^*): \|\theta\|_2 \geq r\}$. Based on this definition of $\Theta_1(r)$, we put

$$\mathcal{R}(r) = \inf_{\tilde{\Delta}} \left\{ \mathbf{P}_0(\tilde{\Delta} = 1) + \sup_{\theta \in \Theta_1(r)} \mathbf{P}_\theta(\tilde{\Delta} = 0) \right\}.$$

The central quantity of our interest is the *minimax testing rate*.

Definition 2.3. The quantity $\varphi_{\varepsilon,\sigma}^2 > 0$ is called *minimax testing rate* if the following two conditions are fulfilled:

(i) for any $\delta \in (0, 1)$, there exists $C^* > 0$ such that for all $C > C^*$ it holds

$$\mathcal{R}(C\varphi_{\varepsilon,\sigma}) \leq \delta,$$

(ii) for any $\delta \in (0, 1)$, there exists $C_* > 0$ such that for all $0 < c < C_*$ it holds

$$\mathcal{R}(c\varphi_{\varepsilon,\sigma}) \geq 1 - \delta.$$

Given this purely non-asymptotic definition, the strategy for deriving the minimax testing rate is as follows: in order to prove the upper bound given by Condition (i), one takes an arbitrary $\delta > 0$ and proposes a test statistic $\widehat{\Delta}$ satisfying

$$\mathbf{P}_0(\widehat{\Delta} = 1) + \sup_{\theta \in \Theta_1(C\varphi_{\varepsilon,\sigma})} \mathbf{P}_1(\widehat{\Delta} = 0) \leq \delta$$

for all C sufficiently large. The proof of the lower bound (ii) is similar to the one of lower bounds for the estimation problem and is mainly based on the auxiliary Lemma A.2 in the appendix. This two-step program is realized for signal detection ($\theta^\circ = 0$) and goodness-of-fit testing ($\theta^\circ \neq 0$) separately in Section 5.

3. MINIMAX UPPER BOUND

Our first theorem provides an upper risk bound for the estimator $\widehat{\mathbf{q}}_k$ for arbitrary $k \in \mathbb{N}^*$.

Theorem 3.1. *Let Assumption 2.2 hold. Then, for any $k \in \mathbb{N}^*$, the estimator $\widehat{\mathbf{q}}_k$ defined in (2.1) satisfies, for any $\theta, \theta^\circ \in \Theta$, the risk bound*

$$\begin{aligned} \sup_{\lambda \in \mathcal{E}} \mathbf{E}[(\widehat{\mathbf{q}}_k - \mathbf{q}(\theta))]^2 &\leq C(d)\varepsilon^4 \sum_{j=1}^k \omega_j^4 \alpha_j^{-4} + C(d)\sigma^4 \sum_{j=1}^k \omega_j^4 \alpha_j^{-4} (\theta_j^\circ)^4 \\ &+ C(d)\varepsilon^2 \sum_{j=1}^k \omega_j^4 \alpha_j^{-2} (\theta_j - \theta_j^\circ)^2 + C(d)\sigma^2 \sum_{j=1}^k \omega_j^4 \alpha_j^{-2} (\theta_j^\circ)^2 (\theta_j - \theta_j^\circ)^2 \\ &+ C(d, L)\sigma^4 \sum_{j=1}^k \omega_j^4 \alpha_j^{-4} \gamma_j^{-2} (\theta_j - \theta_j^\circ)^2 + C(d, L)\sigma^2 \sum_{j=1}^k \omega_j^4 \alpha_j^{-2} \gamma_j^{-2} (\theta_j - \theta_j^\circ)^2 \\ &+ \frac{\omega_k^4}{\gamma_k^4} \left(\sum_{j>k} \gamma_j^2 (\theta_j - \theta_j^\circ)^2 \right)^2. \end{aligned}$$

Consequently, for $\theta^\circ \in \Theta$,

$$\begin{aligned} \sup_{\theta \in \Theta} \sup_{\lambda \in \mathcal{E}} \mathbf{E}[(\widehat{\mathbf{q}}_k - \mathbf{q}(\theta))]^2 &\leq C(d)\varepsilon^4 \sum_{j=1}^k \omega_j^4 \alpha_j^{-4} + C(d, L)\varepsilon^2 \max_{j \in [1, k]} \frac{\omega_j^4}{\alpha_j^2 \gamma_j^2} + C(L) \frac{\omega_k^4}{\gamma_k^4} \\ &+ C(d, L)\sigma^2 \max_{j \in [1, k]} \frac{\omega_j^4}{\alpha_j^2 \gamma_j^4} + C(d, L)\sigma^4 \max_{j \in [1, k]} \frac{\omega_j^4}{\alpha_j^4 \gamma_j^4}. \end{aligned}$$

Proof. We consider the decomposition $\widehat{\mathbf{q}}_k - \mathbf{q}(\theta) = \mathcal{T}_{k1} + \mathcal{T}_{k2} + \mathcal{T}_{k3} + \mathcal{T}_{k4}$ where

$$\begin{aligned}\mathcal{T}_{k1} &= \sum_{j=1}^k \omega_j^2 \frac{U_j}{V_j} \mathbb{1}_{\Omega_j} - \sum_{j=1}^k \omega_j^2 \frac{\lambda_j^2 (\theta_j - \theta_j^\circ)^2}{V_j} \mathbb{1}_{\Omega_j}, \\ \mathcal{T}_{k2} &= \sum_{j=1}^k \omega_j^2 \frac{\lambda_j^2 (\theta_j - \theta_j^\circ)^2}{V_j} \mathbb{1}_{\Omega_j} - \sum_{j=1}^k \omega_j^2 (\theta_j - \theta_j^\circ)^2 \mathbb{1}_{\Omega_j}, \\ \mathcal{T}_{k3} &= - \sum_{j=1}^k \omega_j^2 (\theta_j - \theta_j^\circ)^2 \mathbb{1}_{\Omega_j^c}, \\ \mathcal{T}_{k4} &= \sum_{j>k} \omega_j^2 (\theta_j - \theta_j^\circ)^2.\end{aligned}$$

Thus $\mathbf{E}[(\widehat{\mathbf{q}}_k - \mathbf{q}(\theta))^2] \leq 4 \sum_{i=1}^4 \mathbf{E}\mathcal{T}_{ki}^2$, and the rest of the proof consists in finding appropriate upper bounds for the terms $\mathbf{E}\mathcal{T}_{ki}^2$, $i \in \llbracket 1, 4 \rrbracket$, which are derived in Appendix B. \square

The upper bound proved in Theorem 3.1 consists of terms that are non-decreasing in k , and the term $\omega_k^4 \gamma_k^{-4}$ which is non-increasing in k . Putting

$$k_\varepsilon = \operatorname{argmin}_{k \in \mathbb{N}^*} \max \left\{ \varepsilon^4 \sum_{j=1}^k \omega_j^4 \alpha_j^{-4}, \varepsilon^2 \max_{j \in \llbracket 1, k \rrbracket} \frac{\omega_j^4}{\alpha_j^2 \gamma_j^2}, \frac{\omega_k^4}{\gamma_k^4} \right\} \quad (3.1)$$

and

$$k_\sigma = \operatorname{argmin}_{k \in \mathbb{N}^*} \max \left\{ \sigma^2 \max_{j \in \llbracket 1, k \rrbracket} \frac{\omega_j^4}{\alpha_j^2 \gamma_j^4}, \sigma^4 \max_{j \in \llbracket 1, k \rrbracket} \frac{\omega_j^4}{\alpha_j^4 \gamma_j^4}, \frac{\omega_k^4}{\gamma_k^4} \right\}, \quad (3.2)$$

the quantity k_ε yields the best balance between the squared bias and the variance terms in ε , and analogously k_σ the best balance between squared bias and variance terms in terms of σ . Thus, the following corollary holds.

Corollary 3.2. *Under the assumptions of Theorem 3.1, $k^* := k_\varepsilon \wedge k_\sigma$ with k_ε, k_σ as in (3.1) and (3.2) provides the optimal choice of k in Theorem 3.1, and it holds*

$$\begin{aligned}\sup_{\theta \in \Theta} \sup_{\lambda \in \mathcal{E}} \mathbf{E}[(\widehat{\mathbf{q}}_{k^*} - \mathbf{q}(\theta))^2] &\lesssim \varepsilon^4 \sum_{j=1}^{k^*} \omega_j^4 \alpha_j^{-4} + \varepsilon^2 \max_{j \in \llbracket 1, k^* \rrbracket} \frac{\omega_j^4}{\alpha_j^2 \gamma_j^2} + \frac{\omega_{k^*}^4}{\gamma_{k^*}^4} \\ &\quad + \sigma^2 \max_{j \in \llbracket 1, k^* \rrbracket} \frac{\omega_j^4}{\alpha_j^2 \gamma_j^4} + \sigma^4 \max_{j \in \llbracket 1, k^* \rrbracket} \frac{\omega_j^4}{\alpha_j^4 \gamma_j^4}\end{aligned}$$

where the numerical constant in \lesssim depends on d and L .

It is remarkable that for the estimation of quadratic functionals the optimal truncation parameter k^* depends on both ε and σ whereas the optimal truncation parameter for the estimation of θ itself under ℓ^2 -loss can be chosen in dependence on ε only (see [26], Thm. 2.5). It is not difficult to obtain the rates of convergence for the specific choices of γ and α introduced in Section 2.3 (see Appendix D for some detailed calculations). These rates are summarized in Table 1 for the special case that $\omega_j \equiv 1$. Note that in all illustrations the rate in σ is at least as fast as the one in ε , a fact that can in general be seen from the abstract rates in Corollary 3.2. In some examples, the rate in σ is even strictly faster than the one in ε . For instance in the case where all

the smoothness assumptions are polynomial, one has, depending on the actual values of p and a , to distinguish between three ‘zones’ of rates:

1. if $2p \leq a$, then both rates are non-parametric and the overall rate is $\varepsilon^{16p/(4a+4p+1)} \vee \sigma^{4p/a}$,
2. if $2p \geq a$ but $p \leq a + 1/4$, then the rate in σ is the parametric rate but with respect to ε we are in the non-parametric regime, and the overall rate is $\varepsilon^{16p/(4a+4p+1)} \vee \sigma^2$,
3. if $p \geq a + 1/4$, then we are in the parametric regime with respect to both noise levels and the rate is $\varepsilon^2 \vee \sigma^2 = (\varepsilon \vee \sigma)^2$.

This behaviour generalizes the classical elbow effect which is well known in terms of the noise level ε .

Remark 3.3. Let us mention that, using estimates similar to the ones used in the proof of Theorem 3.1, it would be possible to show that the estimator

$$\widehat{q}_k = \sum_{j=1}^k \frac{X_j^2 - \varepsilon^2}{Y_j^2 - \sigma^2} \mathbf{1}_{\{Y_j^2 \geq 2\sigma^2\}} - 2 \sum_{j=1}^k \theta_j^\circ \frac{X_j}{Y_j} \mathbf{1}_{\{Y_j^2 \geq 2\sigma^2\}} + \sum_{j=1}^k (\theta_j^\circ)^2$$

attains the optimal rate of convergence provided that the truncation parameter is suitably chosen. Note that this estimator does not depend on the availability of two independent samples of the noisy eigenvalues. A theoretical guarantee similar to the one of Theorem 3.1 could, for this alternative estimator, be obtained under an even weaker assumption than Gaussianity (for instance, sub-Gaussianity) because no sample splitting is necessary for the definition of the estimator. However, we stick to the estimator defined in (2.1) since it provides a representation of the risk bound that is more convenient for our application to testing. More precisely, several terms in the risk bound contain the expression $(\theta_j - \theta_j^\circ)^2$ which vanishes when $\theta = \theta^\circ$, and precisely this property is exploited when controlling the type I error of our test procedures.

4. MINIMAX LOWER BOUNDS

In this section, we derive lower bounds on the minimax risk in the sense of (2.2). In order to cleanse the notation, we restrict ourselves without loss of generality to the case $\theta^\circ = 0$ (the proofs in the general case follow easily by adapting this proof to the case $\theta^\circ \neq 0$). Note that the assumptions imposed in addition to Assumption 2.2 in this section are satisfied by all our illustrating examples. Thus, the results of this section imply the optimality of the rates in Table 1.

4.1. Lower bounds in terms of ε

The following theorem provides a lower bound for the case that the rate with respect to ε is determined by the term $\varepsilon^4 \sum_{j=1}^{\kappa} \omega_j^4 \alpha_j^{-4}$ (non-parametric regime) where

$$\kappa = \operatorname{argmin}_{k \in \mathbb{N}^*} \max \left\{ \varepsilon^4 \sum_{j=1}^k \omega_j^4 \alpha_j^{-4}, \frac{\omega_k^4}{\gamma_k^4} \right\}. \quad (4.1)$$

TABLE 1. Optimal rates of convergence for the estimation of quadratic functionals in case that $\omega_j \equiv 1$. Upper bounds are proved in Section 3, lower bounds in Section 4.

	Sobolev class ($\gamma_j = j^p$)	Analytic class ($\gamma_j \asymp e^{pj}$)
Mildly ill-posed ($\alpha_j = j^{-a}$)	$\varepsilon^{16p/(4a+4p+1)} \vee \varepsilon^2 \vee \sigma^{4p/a} \vee \sigma^2$	$\varepsilon^2 \vee \sigma^2$
Severely ill-posed ($\alpha_j \asymp e^{-aj}$)	$ \log \varepsilon ^{-4p} \vee \log \sigma ^{-4p}$	$\varepsilon^{4p/(p+a)} \vee \varepsilon^2 \vee \sigma^{4p/a} \vee \sigma^2$

Theorem 4.1. *Let Assumption 2.2 hold true, and let κ be defined as in (4.1). If*

$$\varepsilon^4 \sum_{j=1}^{\kappa} \omega_j^4 \alpha_j^{-4} \asymp_{\nu} \omega_{\kappa}^4 \gamma_{\kappa}^{-4}$$

for some $\nu \geq 1$, then

$$\inf_{\tilde{\mathbf{q}}} \sup_{\theta \in \Theta} \sup_{\lambda \in \mathcal{E}} \mathbf{E}[(\tilde{\mathbf{q}} - \mathbf{q}(\theta))^2] \gtrsim \varepsilon^4 \sum_{j=1}^{\kappa} \omega_j^4 \alpha_j^{-4}$$

where the infimum is taken over all estimators $\tilde{\mathbf{q}}$ based on the observations (1.1) and (1.2).

The next theorem considers the case that the rate in ε is determined by balancing the terms $\varepsilon^2 \max_{j \in \llbracket 1, k \rrbracket} \omega_j^4 / (\alpha_j \gamma_j)^2$ and the squared bias $\omega_k^4 \gamma_k^{-4}$ (which might result in the parametric rate ε^2 as the lower bound).

Theorem 4.2. *Let Assumption 2.2 hold true.*

(a) *Set*

$$\kappa = \operatorname{argmin}_{k \in \mathbb{N}^*} \max \left\{ \varepsilon^2 \frac{\omega_k^4}{\alpha_k^2 \gamma_k^2}, \frac{\omega_k^4}{\gamma_k^4} \right\},$$

and assume that $\varepsilon^2 \alpha_{\kappa}^{-2} \gamma_{\kappa}^{-2} \asymp_{\nu} \gamma_{\kappa}^{-4}$. Then

$$\inf_{\tilde{\mathbf{q}}} \sup_{\theta \in \Theta} \sup_{\lambda \in \mathcal{E}} \mathbf{E}[(\tilde{\mathbf{q}} - \mathbf{q}(\theta))^2] \gtrsim \varepsilon^2 \frac{\omega_{\kappa}^4}{\alpha_{\kappa}^2 \gamma_{\kappa}^2}$$

where the infimum is taken over all estimators $\tilde{\mathbf{q}}$.

(b) *It holds*

$$\inf_{\tilde{\mathbf{q}}} \sup_{\theta \in \Theta} \sup_{\lambda \in \mathcal{E}} \mathbf{E}[(\tilde{\mathbf{q}} - \mathbf{q}(\theta))^2] \geq \min \left\{ \frac{L^4}{16}, \frac{1}{4d^4} \right\} \cdot \varepsilon^2$$

where the infimum is taken over all estimators $\tilde{\mathbf{q}}$.

For the illustrative examples of rates in Table 1, two different scenarios can occur. In the first case, the sequence $\omega^4 \alpha^{-2} \gamma^{-2}$ is eventually monotone and tends to infinity. Then, statement (a) of Theorem 4.2 applies with the additional assumption of this statement being fulfilled for all our examples. In the second case, the sequence $\omega^4 \alpha^{-2} \gamma^{-2}$ is bounded from above, and we apply statement (b) in order to get the parametric rate ε^2 as a lower bound. The proof of the parametric rate ε^2 in (b) given in Appendix E.2 might be of independent interest, since it provides an alternative to the classical approach given in [17] (see also [18]) who reduces the proof of the lower bound ε^2 to the estimation of a quadratic functional in the normal bounded mean model.

4.2. Lower bounds in terms of σ

We now tackle the problem of finding lower bounds with respect to the noise level σ .

Theorem 4.3. *Let Assumption 2.2 hold true.*

(a) *Set*

$$\kappa = \operatorname{argmin}_{k \in \mathbb{N}^*} \frac{\omega_k^4}{\gamma_k^4} \max\{\sigma^2 \alpha_k^{-2}, 1\},$$

and assume $\sigma^2 \alpha_\kappa^{-2} \asymp_\nu 1$ for some $\nu \geq 1$ independent of σ . Then,

$$\inf_{\tilde{q}} \sup_{\theta \in \Theta} \sup_{\lambda \in \mathcal{E}} \mathbf{E}[(\tilde{q} - \mathbf{q}(\theta))^2] \gtrsim \min_{k \in \mathbb{N}^*} \omega_k^4 \gamma_k^{-4} \max\{\sigma^2 \alpha_k^{-2}, 1\}$$

where the infimum is taken over all estimators \tilde{q} of $\mathbf{q}(\theta)$.

(b) *It holds*

$$\inf_{\tilde{q}} \sup_{\theta \in \Theta} \sup_{\lambda \in \mathcal{E}} \mathbf{E}[(\tilde{q} - \mathbf{q}(\theta))^2] \gtrsim \sigma^2$$

where the infimum is taken over all estimators \tilde{q} of $\mathbf{q}(\theta)$.

As for Theorem 4.2, essentially two scenarios can occur. In the first case, the sequence $\omega^4 \alpha^{-2} \gamma^{-4}$ is eventually monotone and tends to infinity. Then, statement (a) of Theorem 4.3 gives the desired lower bound and the additional assumption in statement (a) is fulfilled by all our examples. In the second case, the sequence $\omega^4 \alpha^{-2} \gamma^{-4}$ is bounded from above, and we apply statement (b) in order to get the parametric rate σ^2 as a lower bound.

5. APPLICATION TO TESTING PROBLEMS

As announced in the introduction we apply the theory developed in the previous sections to signal detection and goodness-of-fit testing separately.

5.1. Signal detection

We start by considering the signal detection problem of testing

$$\mathcal{H}_0 : \theta = 0 \quad \text{against} \quad \mathcal{H}_1 : \theta \in \Theta_1(r)$$

for $r > 0$ where $\Theta_1(r) = \Theta \cap \{\theta : \|\theta\|_2 \geq r\}$. It turns out that for this problem it is sufficient to consider observations (1.1), and to construct a test statistic which is based on an estimator of the quadratic functional $\mathbf{q}^{\text{sd}}(\tilde{\theta}) = \sum_{j=1}^\infty \alpha_j^{-2} \tilde{\theta}_j^2$ where $\tilde{\theta} = \lambda \theta$. Note that the estimation of this quadratic functional from (1.1) is not an inverse but a direct problem since, in terms of $\tilde{\theta}$, (1.1) reads

$$X_j = \tilde{\theta}_j + \varepsilon \xi_j. \tag{5.1}$$

Moreover, the smoothness assumptions in the original model transfer to smoothness assumptions for $\tilde{\theta}$, namely that $\tilde{\theta}$ belongs to an ellipsoid with weight sequence $\tilde{\gamma} = \gamma \alpha^{-1}$. In addition, the weight sequence in the definition $\mathbf{q}^{\text{sd}}(\tilde{\theta})$ is $\omega_j = \alpha_j^{-1}$ in this case. The choice of the truncation value for our auxiliary estimator is slightly different from the optimal choice in Corollary 3.2. More precisely, we put

$$\kappa_1 = \operatorname{argmin}_{k \in \mathbb{N}^*} \max \left\{ \varepsilon^4 \sum_{j=1}^k \alpha_j^{-4}, \gamma_k^{-4} \right\},$$

TABLE 2. Optimal minimax rates of testing for the signal detection problem under the assumptions of Theorems 5.1 and 5.2.

	Sobolev class ($\gamma_j = j^p$)	Analytic class ($\gamma_j \asymp e^{pj}$)
Mildly ill-posed ($\alpha_j = j^{-a}$)	$\varepsilon^{\frac{8p}{4a+4p+1}}$	$\varepsilon^2 \log \varepsilon ^{2a+\frac{1}{2}}$
Severely ill-posed ($\alpha_j \asymp e^{-aj}$)	$ \log \varepsilon ^{-2p}$	$\varepsilon^{\frac{2p}{a+p}}$

and define

$$\hat{q}_{\kappa_1} = \sum_{j=1}^{\kappa_1} \alpha_j^{-2} (X_j^2 - \varepsilon^2). \tag{5.2}$$

Now, in order to prove the upper bound for the minimax testing rate, introduce the test statistic

$$\hat{\Delta}^{\text{sd}} = \mathbb{1}_{\{\hat{q}_{\kappa_1} \geq \tilde{C}\varphi_\varepsilon^2\}} \quad \text{where} \quad \varphi_\varepsilon^2 = \varepsilon^2 \sqrt{\sum_{j=1}^{\kappa_1} \alpha_j^{-4}},$$

and \tilde{C} is a numerical constant that has to be chosen appropriately, see Theorem 5.1. The proof of the following Theorem 5.1 shows that the test statistic $\hat{\Delta}^{\text{sd}}$ satisfies property (i) in Definition 2.3 for the rate φ_ε^2 .

Theorem 5.1. *Let Assumption 2.2 be satisfied, and assume that in addition $\gamma_{\kappa_1}^{-2} \leq \sqrt{\nu}\varphi_\varepsilon^2$ for some $\nu \geq 1$. Let $\delta \in (0, 1)$ be fixed. Then, $\mathcal{R}(C\varphi_\varepsilon) \leq \delta$ for all sufficiently large C .*

The next theorem provides the corresponding lower bound in the sense of condition (ii) from Definition 2.3.

Theorem 5.2. *Let $\delta \in (0, 1)$ be arbitrary. Let Assumption 2.2 hold true, and assume in addition that $\varepsilon^4 \sum_{j=1}^{\kappa_1} \alpha_j^{-4} \asymp_\nu \gamma_{\kappa_1}^{-4}$. Then, there exists $C_* > 0$ such that for all $0 < c < C_*$ the inequality $\mathcal{R}(c\varphi_\varepsilon) \geq 1 - \delta$ holds.*

Specializing the results of Theorems 5.1 and 5.2 with our standard illustrations, we obtain the minimax rates of testing for signal detection in all the considered cases. These are summarized in Table 2.

5.2. Goodness-of-fit testing

In this subsection, we consider the goodness-of-fit testing problem given by testing

$$\mathcal{H}_0: \theta = \theta^\circ \in \Theta \quad \text{against} \quad \mathcal{H}_1: \theta \in \Theta, \theta - \theta^\circ \in \Theta_1(r).$$

In contrast to the signal detection problem considered above, the minimax rate of testing will now depend also on the noise level σ . In the sequel, we make the technical assumption that *all* the components of θ° are non-zero: if this is not the case, one applies the signal detection methodology from Section 5.1 to test the components θ_j° where $\theta_j^\circ = 0$ and combines this approach with the results derived in the sequel. The fact that θ_j° is non-zero is explicitly exploited in the Theorem 5.5 where it guarantees that the constructed hypotheses concerning the solution θ are distinct. We consider the estimator \hat{q}_{κ_2} of the quadratic functional $q^{\text{gof}}(\theta) = \sum_{j=1}^\infty (\theta_j - \theta_j^\circ)^2$ (that is, $\omega_j \equiv 1$ in terms of our general notation) defined through

$$\hat{q}_{\kappa_2} = \sum_{j=1}^{\kappa_2} \frac{U_j}{V_j} \mathbb{1}_{\Omega_j}$$

TABLE 3. Optimal minimax rates of testing for goodness-of-fit testing under the assumptions of Theorems 5.3 and 5.5.

	Sobolev class ($\gamma_j = j^p$)	Analytic class ($\gamma_j \asymp e^{pj}$)
Mildly ill-posed ($\alpha_j = j^{-a}$)	$\varepsilon^{\frac{8p}{4a+4p+1}} \vee \sigma^2 \vee \sigma^{\frac{2p}{a}}$	$\varepsilon^2 \log \varepsilon ^{2a+\frac{1}{2}} \vee \sigma^2$
Severely ill-posed ($\alpha_j \asymp e^{-aj}$)	$ \log \varepsilon ^{-2p} \vee \log \sigma ^{-2p}$	$\varepsilon^{\frac{2p}{a+p}} \vee \sigma^2 \vee \sigma^{\frac{2p}{a}}$

with U_j, V_j, Ω_j defined as in Section 2.2, and κ_2 defined as

$$\kappa_2 = \operatorname{argmin}_{k \in \mathbb{N}^*} \max \left\{ \varepsilon^2 \sqrt{\sum_{j=1}^k \alpha_j^{-4}}, \sigma^2 \max_{j \in \llbracket 1, k \rrbracket} \alpha_j^{-2} \gamma_j^{-2}, \gamma_k^{-2} \right\}$$

(again the definition of the threshold κ_2 slightly differs from the one in Cor. 3.2). Let us introduce the test statistic

$$\widehat{\Delta}^{\text{gof}} = \mathbf{1}_{\{\widehat{q}_{\kappa_2} \geq \widetilde{C} \varphi_{\varepsilon, \sigma}^2\}} \quad \text{where} \quad \varphi_{\varepsilon, \sigma}^2 = \max \left\{ \varepsilon^2 \sqrt{\sum_{j=1}^{\kappa_2} \alpha_j^{-4}}, \sigma^2 \max_{j \in \llbracket 1, \kappa_2 \rrbracket} \alpha_j^{-2} \gamma_j^{-2} \right\}. \quad (5.3)$$

The following theorem shows that this statistic satisfies the upper bound condition (ii) for \widetilde{C} suitably chosen.

Theorem 5.3. *Let Assumption 2.2 be satisfied, and assume that in addition $\gamma_{\kappa_2}^{-2} \leq \sqrt{\nu} \varphi_{\varepsilon, \sigma}^2$ for some $\nu \geq 1$. Let $\delta \in (0, 1)$ be fixed. Then, we have $\mathcal{R}(C \varphi_\varepsilon) \leq \delta$ for all sufficiently large C .*

Remark 5.4. Note that, given $\alpha, \beta \in (0, 1]$, following the same arguments as in the proof of Theorem 5.3, we could tune the numerical constant \widetilde{C} in the definition of the test statistic such that $\mathbf{P}_0(\widehat{\Delta}^{\text{gof}} = 1) \leq \alpha$ and $\mathbf{P}_\theta(\widehat{\Delta}^{\text{gof}} = 0) \leq \beta$ for all $\theta \in \Theta_1(C \varphi_{\varepsilon, \sigma})$ with C sufficiently large. This shows that the order of the separation rate in the sense of [35] is $\varphi_{\varepsilon, \sigma}^2$ (this rate was only derived as a lower bound in [35] whereas the upper bound in that paper contains an additional logarithmic factor; however the test statistic considered in [35] is already adaptive with respect to the class \mathcal{E} in the sense that its definition does neither depend on α nor on d). It might be of interest to find out if the extra logarithmic factors appearing in the rate of [35] are optimal in the sense that no adaptive testing procedure can do without these terms.

We now prove the lower bound on the minimax rate of testing for goodness-of-fit testing.

Theorem 5.5. *Let $\delta \in (0, 1)$ be arbitrary. Let Assumption 2.2 hold true, and assume that $\varphi_{\varepsilon, \sigma}^4 \asymp_\nu \gamma_{\kappa_2}^{-4}$. Then, there exists $C_* > 0$ such that for all $0 < c < C_*$ the inequality $\mathcal{R}(c \varphi_\varepsilon) \geq 1 - \delta$ holds.*

Again, specializing the results of Theorems 5.3 and 5.5 with our standard illustrations, we obtain the minimax rates of testing for the goodness-of-fit testing problem for all the considered cases. These are summarized in Table 3.

6. CONCLUSION AND OPEN QUESTIONS

We have considered the minimax optimal estimation of quadratic functionals in the Gaussian sequence model given by (1.1) and (1.2), and applied our theoretical findings to testing problems. In particular, we have derived the minimax rates of estimation and minimax rates of testing under mild assumptions that allow us to deal with the classical examples from the literature. A next step for future research might be to transfer the methodology developed in this paper to deconvolution models with unknown error distribution [12, 25]. Apart from that, the

following problems have not been dealt with in this paper and might be worth of being more closely investigated:

- The optimal estimator of the quadratic functionals is not completely data-driven, and the definition of an adaptive selection rule for the truncation parameter that satisfies some theoretical guarantees is necessary.
- Equally, the problem of adaptive testing has not been discussed. In particular, can standard techniques for adaptive testing in inverse problems as developed in [4] (see also [30]) be transferred to the model with partially unknown operator, and what is the price one has to pay for adaption?
- The general matrix case given by observations

$$X = A\theta + \varepsilon\xi \quad \text{and} \quad Y = A + \sigma\Xi$$

is still open. Note that results for this model might be of interest since it is related to inverse problems like non-parametric instrumental variable regression or functional linear regression where non-diagonal matrices appear in a natural manner.

- Finally, considering inverse problems with sparsity constraints as in [11] might be of interest.

APPENDIX A. GENERAL TOOLS FOR LOWER BOUNDS

A.1 Reduction to comparison with a fuzzy hypothesis

For a probability measure μ on Θ , we put $\mathbf{P}_\mu^X = \int_\Theta \mathbf{P}_\theta^X \mu(d\theta)$. The following lemma reduces the problem of establishing a minimax lower bound on the class Θ to the problem of testing \mathbf{P}_0^X ($\mu = \delta_0$) against some \mathbf{P}_μ^X with $\mu \neq \delta_0$. It is a special case of Theorem 2.15 in [38] (the formulation is mainly borrowed from [11], see Lem. 2 therein).

Lemma A.1. *Let Θ be a subset of $\ell^2(\mathbb{N}^*)$ containing 0. Assume that there exists a probability measure μ on Θ and numbers $\psi > 0$, $\beta > 0$ such that $q(\theta) = 2\psi$ for all $\theta \in \text{supp}(\mu)$ and $\chi^2(\mathbf{P}_\mu^X, \mathbf{P}_0^X) \leq \beta$. Then,*

$$\inf_{\tilde{q}} \sup_{\theta \in \Theta} \sup_{\lambda \in \mathcal{E}} \mathbf{P}_{(\theta, \lambda)}(|\tilde{q} - q(\theta)| \geq \psi) \geq \frac{1}{4} \exp(-\beta)$$

where the infimum is taken over all estimators \tilde{q} .

A.2 Reduction to two hypotheses

For the proofs of Theorems 4.2 and 4.3 we will construct hypotheses $(\theta^\tau, \lambda^\tau) \in \Theta \times \mathcal{E}$ for $\tau \in \{\pm 1\}$ such that the Kullback-Leibler distance between the resulting distributions \mathbf{P}_1 and \mathbf{P}_{-1} of the tuple (X, Y) is bounded by 1. This implies $\rho(\mathbf{P}_1, \mathbf{P}_{-1}) \geq 1/2$ for the Hellinger affinity being defined as $\rho(\mathbf{P}_1, \mathbf{P}_{-1}) = \sqrt{d\mathbf{P}_1 d\mathbf{P}_{-1}}$. Putting $q_\tau = q(\theta^\tau)$ for $\tau \in \{\pm 1\}$ we can conclude from

$$\begin{aligned} \frac{1}{2} &\leq \int \frac{|\tilde{q} - q_1|}{|q_1 - q_{-1}|} \sqrt{d\mathbf{P}_1 d\mathbf{P}_{-1}} + \int \frac{|\tilde{q} - q_{-1}|}{|q_1 - q_{-1}|} \sqrt{d\mathbf{P}_1 d\mathbf{P}_{-1}} \\ &\leq \left(\int \left(\frac{\tilde{q} - q_1}{q_1 - q_{-1}} \right)^2 d\mathbf{P}_1 \right)^{1/2} + \left(\int \left(\frac{\tilde{q} - q_{-1}}{q_1 - q_{-1}} \right)^2 d\mathbf{P}_{-1} \right)^{1/2} \end{aligned}$$

by using the elementary estimate $(a + b)^2 \leq 2a^2 + 2b^2$ that

$$\frac{1}{8} (q_1 - q_{-1})^2 \leq \mathbf{E}_1[(\tilde{q} - q_1)^2] + \mathbf{E}_{-1}[(\tilde{q} - q_{-1})^2].$$

This last estimate in turn yields

$$\sup_{\theta \in \Theta} \sup_{\lambda \in \mathcal{E}} \mathbf{E}[(\tilde{\mathbf{q}} - \mathbf{q}(\theta))^2] \geq \frac{1}{2} \sum_{\tau \in \{\pm 1\}} \mathbf{E}_\tau[(\tilde{\mathbf{q}} - \mathbf{q}_\tau)^2] \geq \frac{1}{16} (\mathbf{q}_1 - \mathbf{q}_{-1})^2 \tag{A.1}$$

which establishes the quantity $\frac{1}{16}(\mathbf{q}_1 - \mathbf{q}_{-1})^2$ as a lower bound on the minimax rate.

A.3 Reduction argument for lower bounds of testing

Lemma A.2. *Let μ be a probability measure on Θ_1 . Then, the following statements hold true:*

- (i) $\inf_{\Delta} \{ \mathbf{P}_0(\Delta = 1) + \sup_{\theta \in \Theta_1} \mathbf{P}_\theta(\Delta = 0) \} \geq 1 - \sqrt{\chi^2(\mathbf{P}_\mu, \mathbf{P}_0)}$,
- (ii) $\inf_{\Delta} \{ \mathbf{P}_0(\Delta = 1) + \sup_{\theta \in \Theta_1} \mathbf{P}_\theta(\Delta = 0) \} \geq 1 - \sqrt{\text{KL}(\mathbf{P}_\mu, \mathbf{P}_0)/2}$.

In both statements, the infimum is taken over all $\{0, 1\}$ -valued statistics.

Proof. For any $\{0, 1\}$ -valued statistic Δ ,

$$\begin{aligned} \mathbf{P}_0(\Delta = 1) + \sup_{\theta \in \Theta_1} \mathbf{P}_\theta(\Delta = 0) &\geq \mathbf{P}_0(\Delta = 1) + \int_{\Theta_1} \mathbf{P}_\theta(\Delta = 0) \mu(d\theta) \\ &= \mathbf{P}_0(\Delta = 1) + \mathbf{P}_\mu(\Delta = 0) \\ &\geq 1 - \text{TV}(\mathbf{P}_\mu, \mathbf{P}_0). \end{aligned}$$

Therefrom, statement (i) follows using equation (2.27) in [38], and statement (ii) by the first Pinsker inequality (see [38], Lem. 2.5). \square

APPENDIX B. UPPER BOUNDS FOR THE TERMS $\mathbf{E}\mathcal{T}_{ki}^2$ IN THE PROOF OF THEOREM 3.1

Upper bound for $\mathbf{E}\mathcal{T}_{k1}^2$. By independence of U_j and $\mathbb{1}_{\Omega_j}/V_j$ and $\mathbf{E}[U_j - \lambda_j^2(\theta_j - \theta_j^\circ)] = 0$, it holds

$$\begin{aligned} \mathbf{E} \left[\left(\sum_{j=1}^k \omega_j^2 \frac{U_j - \lambda_j^2(\theta_j - \theta_j^\circ)^2}{V_j} \mathbb{1}_{\Omega_j} \right)^2 \right] &= \text{Var} \left(\sum_{j=1}^k \omega_j^2 \frac{U_j - \lambda_j^2(\theta_j - \theta_j^\circ)^2}{V_j} \mathbb{1}_{\Omega_j} \right) \\ &= \sum_{j=1}^k \omega_j^4 \text{Var} \left(\frac{U_j - \lambda_j^2(\theta_j - \theta_j^\circ)^2}{V_j} \mathbb{1}_{\Omega_j} \right). \end{aligned}$$

Set $Z_1 = (U_j - \lambda_j^2(\theta_j - \theta_j^\circ)^2)/\lambda_j^2$ and $Z_2 = \lambda_j^2/V_j \cdot \mathbb{1}_{\Omega_j}$. Note that Z_1 and Z_2 are independent, and since $\mathbf{E}Z_1 = 0$, the identity $\text{Var}(Z_1 Z_2) = \text{Var}(Z_1) \text{Var}(Z_2) + \text{Var}(Z_1)(\mathbf{E}Z_2)^2 + \text{Var}(Z_2)(\mathbf{E}Z_1)^2$ reduces to $\text{Var}(Z_1 Z_2) = \text{Var}(Z_1)\mathbf{E}(Z_2^2)$. Hence,

$$\begin{aligned} \text{Var} \left(\frac{U_j - \lambda_j^2(\theta_j - \theta_j^\circ)^2}{V_j} \mathbb{1}_{\Omega_j} \right) &= \text{Var}(Z_1)\mathbf{E}[Z_2^2] \\ &= \text{Var} \left(\frac{U_j - \lambda_j^2(\theta_j - \theta_j^\circ)^2}{\lambda_j^2} \right) \cdot \mathbf{E} \left[\frac{\lambda_j^4}{V_j^2} \mathbb{1}_{\Omega_j} \right] \\ &\leq 168\lambda_j^{-4} \text{Var}(U_j) \end{aligned}$$

where we have used statement (i) from Proposition C.1. Now, since $\text{Var}(U_j) = 2(\varepsilon^2 + 2\sigma^2(\theta_j^\circ)^2)^2 + 4(\varepsilon^2 + 2\sigma^2(\theta_j^\circ)^2)\lambda_j^2(\theta_j - \theta_j^\circ)^2$, we obtain using $(a + b)^2 \leq 2a^2 + 2b^2$ that

$$\begin{aligned} \text{Var} \left(\frac{U_j - \lambda_j^2(\theta_j - \theta_j^\circ)^2}{V_j} \mathbf{1}_{\Omega_j} \right) \\ \leq 672\varepsilon^4\lambda_j^{-4} + 2688\sigma^4(\theta_j^\circ)^4\lambda_j^{-4} + 672(\varepsilon^2 + 2\sigma^2(\theta_j^\circ)^2)\lambda_j^{-2}(\theta_j - \theta_j^\circ)^2. \end{aligned}$$

Now summation over all indices $j \in \llbracket 1, k \rrbracket$ yields

$$\begin{aligned} \mathbf{E}\mathcal{T}_{k1}^2 &\leq 672\varepsilon^4 \sum_{j=1}^k \omega_j^4 \lambda_j^{-4} + 2688\sigma^4 \sum_{j=1}^k \omega_j^4 \lambda_j^{-4} (\theta_j^\circ)^4 \\ &\quad + 672\varepsilon^2 \sum_{j=1}^k \omega_j^4 \lambda_j^{-2} (\theta_j - \theta_j^\circ)^2 + 1344\sigma^2 \sum_{j=1}^k \omega_j^4 \lambda_j^{-2} (\theta_j^\circ)^2 (\theta_j - \theta_j^\circ)^2 \\ &\leq 672d^4\varepsilon^4 \sum_{j=1}^k \omega_j^4 \alpha_j^{-4} + 2688d^4\sigma^4 \sum_{j=1}^k \omega_j^4 \alpha_j^{-4} (\theta_j^\circ)^4 \\ &\quad + 672d^2\varepsilon^2 \sum_{j=1}^k \omega_j^4 \alpha_j^{-2} (\theta_j - \theta_j^\circ)^2 + 1344d^2\sigma^2 \sum_{j=1}^k \omega_j^4 \alpha_j^{-2} (\theta_j^\circ)^2 (\theta_j - \theta_j^\circ)^2. \end{aligned}$$

Upper bound for $\mathbf{E}\mathcal{T}_{k2}^2$. Using the Cauchy-Schwarz inequality, it holds

$$\begin{aligned} \mathbf{E}\mathcal{T}_{2k}^2 &= \mathbf{E} \left[\left(\sum_{j=1}^k \omega_j^2 \lambda_j^2 (\theta_j - \theta_j^\circ)^2 \left(\frac{1}{V_j} - \frac{1}{\lambda_j^2} \right) \mathbf{1}_{\Omega_j} \right)^2 \right] \\ &\leq \mathbf{E} \left[\left(\sum_{j=1}^k \gamma_j^2 (\theta_j - \theta_j^\circ)^2 \right) \left(\sum_{j=1}^k \omega_j^4 \gamma_j^{-2} (\theta_j - \theta_j^\circ)^2 \left(\frac{\lambda_j^2}{V_j} - 1 \right)^2 \mathbf{1}_{\Omega_j} \right) \right] \\ &\leq 4L^2 \sum_{j=1}^k \omega_j^4 (\theta_j - \theta_j^\circ)^2 \gamma_j^{-2} \mathbf{E} \left[\left(\frac{\lambda_j^2}{V_j} - 1 \right)^2 \mathbf{1}_{\Omega_j} \right] \\ &\leq C(d)L^2 \sum_{j=1}^k \omega_j^4 \gamma_j^{-2} (\sigma^4 \alpha_j^{-4} + \sigma^2 \alpha_j^{-2}) (\theta_j - \theta_j^\circ)^2 \end{aligned}$$

where the last estimate is due to statement (ii) from Proposition C.1.

Upper bound for $\mathbf{E}\mathcal{T}_{k3}^2$. Again by the Cauchy-Schwarz inequality we have

$$\begin{aligned} \mathbf{E} \left[\left(\sum_{j=1}^k \omega_j^2 (\theta_j - \theta_j^\circ)^2 \mathbf{1}_{\Omega_j^c} \right)^2 \right] &\leq \mathbf{E} \left[\left(\sum_{j=1}^k \gamma_j^2 (\theta_j - \theta_j^\circ)^2 \right) \left(\sum_{j=1}^k \omega_j^4 \gamma_j^{-2} (\theta_j - \theta_j^\circ)^2 \mathbf{1}_{\Omega_j^c} \right) \right] \\ &\leq 4L^2 \sum_{j=1}^k \omega_j^4 \gamma_j^{-2} (\theta_j - \theta_j^\circ)^2 \mathbf{P}(\Omega_j^c). \end{aligned}$$

Bounding the probability of the event Ω_j^c by means of statement (iii) in Proposition C.1, we conclude

$$\mathbf{E}\mathcal{T}_{k3}^2 \leq 48d^2L^2 \sum_{j=1}^k \omega_j^4 \gamma_j^{-2} (\theta_j - \theta_j^\circ)^2 \min\{1, \sigma^2 \alpha_j^{-2}\}.$$

Upper bound for $\mathbf{E}\mathcal{T}_{k4}^2$. Note that \mathcal{T}_{k4}^2 is deterministic. Hence,

$$\mathbf{E}\mathcal{T}_{k4}^2 = \mathcal{T}_{k4}^2 = \left(\sum_{j>k} \omega_j^2 (\theta_j - \theta_j^\circ)^2 \right)^2 \leq \frac{\omega_k^4}{\gamma_k^4} \left(\sum_{j>k} \gamma_j^2 (\theta_j - \theta_j^\circ)^2 \right)^2 \leq 16L^4 \cdot \frac{\omega_k^4}{\gamma_k^4}.$$

APPENDIX C. AUXILIARY RESULTS FOR THE PROOF OF THEOREM 3.1

Proposition C.1. *With the notations introduced in the main part of the paper the following assertions hold true:*

- (i) $\mathbf{E} \left[\frac{\lambda_j^4}{V_j^2} \mathbf{1}_{\Omega_j} \right] \leq 168,$
- (ii) $\mathbf{E} \left[\left(\frac{\lambda_j^2}{V_j} - 1 \right)^2 \mathbf{1}_{\Omega_j} \right] \leq C(d)\sigma^4 \alpha_j^{-4} + C(d)\sigma^2 \alpha_j^{-2},$
- (iii) $\mathbf{P}(\Omega_j^c) \leq 12d^2 \min(1, \sigma^2 \alpha_j^{-2}).$

Proof. We begin the proof of (i) with the observation that, since the function $x \mapsto \frac{x}{x-2\sigma^2}$ is non-increasing on $[3\sigma^2, \infty),$

$$\frac{\bar{Y}_j^4}{V_j^2} \mathbf{1}_{\Omega_j} = \left(\frac{\bar{Y}_j^2}{V_j} \right)^2 \mathbf{1}_{\Omega_j} = \left(\frac{\bar{Y}_j^2}{\bar{Y}_j^2 - 2\sigma^2} \right)^2 \mathbf{1}_{\Omega_j} \leq \left(\frac{3\sigma^2}{\sigma^2} \right)^2 \leq 9. \tag{C.1}$$

Therefrom, using $(a + b)^4 \leq 8a^4 + 8b^4$

$$\begin{aligned} \mathbf{E} \left[\frac{\lambda_j^4}{V_j^2} \mathbf{1}_{\Omega_j} \right] &\leq \mathbf{E} \left[\frac{\lambda_j^4}{\bar{Y}_j^4} \cdot \frac{\bar{Y}_j^4}{V_j^2} \mathbf{1}_{\Omega_j} \right] \leq 9 \mathbf{E} \left[\frac{\lambda_j^4}{\bar{Y}_j^4} \mathbf{1}_{\Omega_j} \right] \\ &\leq 9 \mathbf{E} \left[\frac{(\lambda_j - \bar{Y}_j + \bar{Y}_j)^4}{\bar{Y}_j^4} \mathbf{1}_{\Omega_j} \right] \\ &\leq 72 \mathbf{E} \left[\frac{(\lambda_j - \bar{Y}_j)^4}{9\sigma^4} \right] + 72 \leq 96 + 72 = 168. \end{aligned}$$

In order to show (ii), introduce the event $\mathcal{U}_j := \left\{ \left| \frac{1}{\bar{Y}_j} - \frac{1}{\lambda_j} \right| \leq \frac{1}{2|\lambda_j|} \right\}.$ Then, trivially,

$$\mathbf{E} \left[\left(\frac{\lambda_j^2}{V_j} - 1 \right)^2 \mathbf{1}_{\Omega_j} \right] = \mathbf{E} \left[\left(\frac{\lambda_j^2}{V_j} - 1 \right)^2 \mathbf{1}_{\Omega_j} (\mathbf{1}_{\mathcal{U}_j} + \mathbf{1}_{\mathcal{U}_j^c}) \right], \tag{C.2}$$

and we consider the summands with $\mathbf{1}_{\mathcal{U}_j}$ and $\mathbf{1}_{\mathcal{U}_j^c}$ separately. First, using (C.1),

$$\mathbf{E} \left[\left(\frac{\lambda_j^2}{V_j} - 1 \right)^2 \mathbf{1}_{\Omega_j} \mathbf{1}_{\mathcal{U}_j} \right] = \mathbf{E} \left[\frac{(\lambda_j^2 - V_j)^2}{\bar{Y}_j^4} \cdot \frac{\bar{Y}_j^4}{V_j^2} \cdot \mathbf{1}_{\Omega_j} \mathbf{1}_{\mathcal{U}_j} \right] \leq 9 \mathbf{E} \left[\frac{(\lambda_j^2 - V_j)^2}{\bar{Y}_j^4} \cdot \mathbf{1}_{\mathcal{U}_j} \right],$$

and since the definition of \mathcal{U}_j implies that $\bar{Y}_j^{-4} \leq \frac{81}{16} \lambda_j^{-4} \leq \frac{81}{16} d^4 \alpha_j^{-4}$, we have

$$\begin{aligned} \mathbf{E} \left[\left(\frac{\lambda_j^2}{V_j} - 1 \right)^2 \mathbf{1}_{\Omega_j} \mathbf{1}_{\mathcal{U}_j} \right] &\leq \frac{729}{16} d^4 \alpha_j^{-4} \mathbf{E}[(\lambda_j^2 - V_j)^2] = \frac{729}{16} d^4 \alpha_j^{-4} (8\sigma^4 + 8\sigma^2 \lambda_j^2) \\ &\leq \frac{729}{2} d^4 \sigma^4 \alpha_j^{-4} + \frac{729}{2} d^6 \sigma^2 \alpha_j^{-2}. \end{aligned}$$

We now turn to the summand with $\mathbf{1}_{\mathcal{U}_j^c}$. First by the Cauchy-Schwarz inequality,

$$\begin{aligned} \mathbf{E} \left[\left(\frac{\lambda_j^2}{V_j} - 1 \right)^2 \mathbf{1}_{\Omega_j} \mathbf{1}_{\mathcal{U}_j^c} \right] &\leq \left(\mathbf{E} \left[\left(\frac{\lambda_j^2}{V_j} - 1 \right)^4 \mathbf{1}_{\Omega_j} \right] \right)^{1/2} \cdot \mathbf{P}(\mathcal{U}_j^c)^{1/2} \\ &\leq \sigma^{-4} \cdot [\mathbf{E}(\lambda_j^2 - V_j)^4]^{1/2} \cdot \mathbf{P}(\mathcal{U}_j^c)^{1/2}. \end{aligned}$$

Now, simple but exhausting calculations show that $\mathbf{E}[(\lambda_j^2 - V_j)^4] = 196\lambda_j^4\sigma^4 + 1920\lambda_j^2\sigma^6 + 960\sigma^8$. Thus, using the estimate $\sqrt{a+b+c} \leq \sqrt{a} + \sqrt{b} + \sqrt{c}$ for $a, b, c \geq 0$, we obtain

$$\begin{aligned} \mathbf{E} \left[\left(\frac{\lambda_j^2}{V_j} - 1 \right)^2 \mathbf{1}_{\Omega_j} \mathbf{1}_{\mathcal{U}_j^c} \right] &\leq \sigma^{-4} (\sqrt{196}\lambda_j^2\sigma^2 + \sqrt{1920}\lambda_j\sigma^3 + \sqrt{960}\sigma^4) \mathbf{P}(\mathcal{U}_j^c)^{1/2} \\ &\leq (\sqrt{196}d^2\alpha_j^2\sigma^{-2} + \sqrt{1920}d\alpha_j\sigma^{-1} + \sqrt{960}) \mathbf{P}(\mathcal{U}_j^c)^{1/2}. \end{aligned}$$

By definition, $\mathcal{U}_j^c = \{|\lambda_j/\bar{Y}_j - 1| > 1/2\}$, which implies that $\lambda_j/\bar{Y}_j \notin [1/2, 3/2]$ on \mathcal{U}_j^c . Hence $\bar{Y}_j/\lambda_j \notin [2/3, 2]$ on \mathcal{U}_j^c showing the inclusion $\mathcal{U}_j^c \subseteq \{|\bar{Y}_j/\lambda_j - 1| > 1/3\} = \{|\bar{Y}_j - \lambda_j| > |\lambda_j|/3\}$, and hence

$$\mathbf{P}(\mathcal{U}_j^c) \leq 2 \exp(-\lambda_j^2/(36\sigma^2)) \leq 2 \exp(-\alpha_j^2/(36d^2\sigma^2)).$$

We obtain

$$\begin{aligned} \mathbf{E} \left[\left(\frac{\lambda_j^2}{V_j} - 1 \right)^2 \mathbf{1}_{\Omega_j} \mathbf{1}_{\mathcal{U}_j^c} \right] &\leq (\sqrt{392}d^2\alpha_j^2\sigma^{-2} + \sqrt{3840}d\alpha_j\sigma^{-1} + \sqrt{1920}) \exp(-\alpha_j^2/(72d^2\sigma^2)). \end{aligned}$$

It is easy to see that there are constants $C_1(d), C_2(d)$ and $C_3(d)$ such that

$$\begin{aligned} \alpha_j^2\sigma^{-2} \exp(-\alpha_j^2/(72d^2\sigma^2)) &\leq C_1(d)\sigma^2\alpha_j^{-2}, \\ \alpha_j\sigma^{-1} \exp(-\alpha_j^2/(72d^2\sigma^2)) &\leq C_2(d)\sigma^2\alpha_j^{-2}, \\ \exp(-\alpha_j^2/(72d^2\sigma^2)) &\leq C_3(d)\sigma^2\alpha_j^{-2}, \end{aligned}$$

and thus

$$\mathbf{E} \left[\left(\frac{\lambda_j^2}{V_j} - 1 \right)^2 \mathbf{1}_{\Omega_j} \mathbf{1}_{V_j^c} \right] \leq C(d) \sigma^2 \alpha_j^{-2}.$$

Now, combining the derived bounds for the two terms on the right hand-side of (C.2) implies the claim assertion. For the proof of (iii), we consider first the case that $\lambda_j^2 \geq 12\sigma^2$. Then, by Chebyshev's inequality,

$$\mathbf{P}(\Omega_j^c) \leq \mathbf{P} \left(\frac{\bar{Y}_j^2}{\lambda_j^2} < \frac{1}{4} \right) \leq \mathbf{P} \left(\left| \frac{\bar{Y}_j}{\lambda_j} - 1 \right| > \frac{1}{2} \right) \leq 8\sigma^2 \lambda_j^{-2} \leq 8d^2 \sigma^2 \alpha_j^{-2}.$$

In case that $\lambda_j^2 \leq 12\sigma^2$, we have $1 \leq 12d^2 \sigma^2 \alpha_j^{-2}$ and $\mathbf{P}(\Omega_j^c) \leq 12d^2 \sigma^2 \alpha_j^{-2}$ holds trivially. Combining the two considered cases implies the claim assertion. \square

APPENDIX D. CALCULATIONS OF RATES

We sketch the calculations leading to the rates in Table 1. Recall that $\omega_j \equiv 1$ for all these examples.

1. Case: $\gamma_j = j^p$, $\alpha_j = j^{-a}$

It holds $\varepsilon^4 \sum_{j=1}^k \alpha_j^{-4} = \varepsilon^4 \sum_{j=1}^k j^{4a} \asymp \varepsilon^4 k^{4a+1}$, and

$$\varepsilon^2 \max_{j \in [1, k]} \alpha_j^{-2} \gamma_j^{-2} = \varepsilon^2 \max_{j \in [1, k]} j^{2a-2p} = \begin{cases} \varepsilon^2, & \text{if } p \geq a, \\ \varepsilon^2 k^{2a-2p}, & \text{if } p < a. \end{cases}$$

Thus, $k_\varepsilon \asymp \varepsilon^{-4/(4a+4p+1)} \wedge \varepsilon^{-1/(2p)}$. Similarly,

$$\sigma^2 \max_{j \in [1, k]} \alpha_j^{-2} \gamma_j^{-4} = \sigma^2 \max_{j \in [1, k]} j^{2a-4p} = \begin{cases} \sigma^2, & \text{if } 2p \geq a, \\ \sigma^2 k^{2a-4p}, & \text{if } 2p < a. \end{cases}$$

Hence, $k_\sigma \asymp \sigma^{-1/a} \wedge \sigma^{-1/(2p)}$. The rate resulting from these values of k_ε and k_σ is

$$\varepsilon^{16p/(4a+4p+1)} \wedge \varepsilon^2 \wedge \sigma^{4p/a} \wedge \sigma^2.$$

2. Case: $\gamma_j \asymp \exp(pj)$, $\alpha_j = j^{-a}$

As in the previous case we have $\varepsilon^4 \sum_{j=1}^k \alpha_j^{-4} \asymp \varepsilon^4 k^{4a+1}$, but now

$$\varepsilon^2 \max_{j \in [1, k]} \alpha_j^{-2} \gamma_j^{-2} \lesssim \varepsilon^2.$$

Balancing the approximation error γ_k^{-4} and ε^2 results in choosing $k_\varepsilon = \lfloor \log \varepsilon / (2p) \rfloor$ which in turn leads to the parametric rate ε^2 . Analogously, $k_\sigma = \lfloor \log \sigma / (2p) \rfloor$, likewise implying the parametric rate σ^2 .

3. Case: $\gamma_j = j^p$, $\alpha_j \asymp \exp(-aj)$

In this case, $\varepsilon^4 \sum_{j=1}^k \alpha_j^{-4} \asymp \varepsilon^4 \sum_{j=1}^k \exp(4ja) \asymp \varepsilon^4 \exp(4ka)$, and balancing this expression with the approximation error γ_k^{-4} leads to a choice of k_ε of order $\log \varepsilon$. Plugging this choice into the approximation error leads, with respect to ε , to the rate $|\log \varepsilon|^{-4p}$. Moreover, in the case at hand,

$$\sigma^2 \max_{j \in [1, k]} \alpha_j^{-2} \gamma_j^{-4} \lesssim \sigma^2 \exp(2ka) \cdot k^{-4p}$$

which implies analogously to the choice of k_ε a choice of k_σ of order $|\log \sigma|$. Hence, the rate with respect to σ is $|\log \sigma|^{-4p}$. Note that the exact knowledge of p and a is not necessary in this case, since it suffices to choose $k_\varepsilon \asymp |\log \varepsilon|$ and $k_\sigma \asymp |\log \sigma|$ (however, one has to know that the unknown solution belongs to a Sobolev class and that the inverse problem is severely ill-posed).

4. Case: $\gamma_j \asymp \exp(pj)$, $\alpha_j \asymp \exp(-aj)$
 We have

$$\begin{aligned} \varepsilon^2 \max_{j \in [1, k]} \alpha_j^{-2} \gamma_j^{-2} &\asymp \varepsilon^2 \max_{j \in [1, k]} \exp(2(a-p)j) \\ &= \begin{cases} \varepsilon^2, & \text{if } p \geq a, \\ \varepsilon^2 \exp(2(a-p)k), & \text{if } p < a. \end{cases} \end{aligned}$$

Hence, $k_\varepsilon = \lfloor |\log \varepsilon| / (a+p) \rfloor \wedge \lfloor |\log \varepsilon| / (2p) \rfloor$ (the second choice would equally originate from balancing the term $\varepsilon^4 \sum_{j=1}^k \alpha_j^{-4}$ with the approximation term γ_k^{-4}). Thus, the resulting rate in terms of ε is $\varepsilon^{4p/(p+a)} \vee \varepsilon^2$. Concerning the rate in terms of σ^2 for this case, note first that

$$\begin{aligned} \sigma^2 \max_{j \in [1, k]} \alpha_j^{-2} \gamma_j^{-4} &\asymp \sigma^2 \max_{j \in [1, k]} \exp(2(a-2p)j) \\ &= \begin{cases} \sigma^2, & \text{if } 2p \geq a, \\ \sigma^2 \exp(2(a-p)k), & \text{if } 2p < a. \end{cases} \end{aligned}$$

Thus, $k_\sigma = \lfloor |\log \sigma| / a \rfloor \wedge \lfloor |\log \sigma| / (2p) \rfloor$ leading to the rate $\sigma^2 \vee \sigma^{4p/a}$ in terms of σ .

APPENDIX E. PROOFS OF SECTION 4

E.1 Proof of Theorem 4.1

By Markov's inequality one has for every estimator \tilde{q} of $q(\theta)$ that

$$\inf_{\tilde{q}} \sup_{\theta \in \Theta} \sup_{\lambda \in \mathcal{E}} \mathbf{E}[(\tilde{q} - q(\theta))^2] \geq \psi^2 \cdot \inf_{\tilde{q}} \sup_{\theta \in \Theta} \sup_{\lambda \in \mathcal{E}} \mathbf{P}((\tilde{q} - q(\theta))^2 \geq \psi^2), \tag{E.1}$$

and we want to apply Lemma A.1 from Appendix A with $\psi = \frac{1}{2} L^2 \nu^{-1/2} \varepsilon^2 \sqrt{\sum_{j=1}^\kappa \omega_j^4 \alpha_j^{-4}}$. For any $\tau = (\tau_1, \dots, \tau_\kappa) \in \{\pm 1\}^\kappa$ define θ^τ via

$$\theta_i^\tau = \tau_i \cdot L \nu^{-1/4} \cdot \varepsilon \cdot \frac{\omega_i \alpha_i^{-2}}{(\sum_{j=1}^\kappa \omega_j^4 \alpha_j^{-4})^{1/4}} \quad \text{for } i \in [1, \kappa],$$

and $\theta_i^\tau = 0$ for $i > \kappa$. Then, for any $\tau \in \{\pm 1\}^\kappa$,

$$\sum_{j=1}^\infty (\theta_j^\tau)^2 \gamma_j^2 = L^2 \nu^{-1/2} \cdot \frac{\varepsilon^2}{(\sum_{i=1}^\kappa \omega_i^4 \alpha_i^{-4})^{1/2}} \sum_{j=1}^\kappa \omega_j^2 \alpha_j^{-4} \gamma_j^2 \leq L^2$$

where we have used that $\varepsilon^2 \omega_j^{-2} \gamma_j^2 \leq \varepsilon^2 \omega_\kappa^{-2} \gamma_\kappa^2 \leq \nu^{1/2} (\sum_{i=1}^\kappa \omega_i^4 \alpha_i^{-4})^{-1/2}$ by assumption. Thus, $\theta^\tau \in \Theta$ for any $\tau \in \{\pm 1\}^\kappa$. Further,

$$q(\theta^\tau) = L^2 \nu^{-1/2} \varepsilon^2 \frac{\sum_{j=1}^\kappa \omega_j^4 / \alpha_j^4}{\sqrt{\sum_{j=1}^\kappa \omega_j^4 / \alpha_j^4}} = L^2 \nu^{-1/2} \varepsilon^2 \sqrt{\sum_{j=1}^\kappa \omega_j^4 / \alpha_j^4}.$$

Consider the probability measure μ on Θ that is induced by the uniform distribution on the hypercube $\{\pm 1\}^\kappa$ via the mapping $\{\pm 1\}^\kappa \rightarrow \Theta$, $\omega \mapsto \theta^\tau$. Let \mathbf{P}_μ be the resulting distribution of the tuple (X, Y) when $\lambda = \lambda^\circ$ for some fixed but arbitrary $\lambda^\circ \in \mathcal{E}$, and analogously \mathbf{P}_0 the distribution of (X, Y) when $\theta = 0 \in \Theta$ and $\lambda = \lambda^\circ$. Computing the χ^2 -distance between \mathbf{P}_μ and \mathbf{P}_0 yields

$$\chi^2(\mathbf{P}_\mu, \mathbf{P}_0) = \int \left(\frac{d\mathbf{P}_\mu}{d\mathbf{P}_0} \right)^2 d\mathbf{P}_0 - 1 = \prod_{j=1}^{\kappa} \frac{\exp(-\lambda_j^2 \beta_j^2 / \varepsilon^2) + \exp(\lambda_j^2 \beta_j^2 / \varepsilon^2)}{2} - 1$$

where $\beta_j = \varepsilon L \nu^{-1/4} \cdot \frac{\omega_j \alpha_j^{-2}}{(\sum_{i=1}^{\kappa} \omega_i^4 \alpha_i^{-4})^{1/4}}$. Now, using the same reasoning as on page 130 in [38], it can be shown that there exists a constant $c_2 < \infty$ such that

$$\frac{\exp(-\lambda_j^2 \beta_j^2 / \varepsilon^2) + \exp(\lambda_j^2 \beta_j^2 / \varepsilon^2)}{2} \leq \exp\left(\frac{c_2 \lambda_j^4 \beta_j^4}{\varepsilon^4}\right).$$

Thus, denoting with c_2 and c_3 numerical constants that depend on d , we conclude

$$\chi^2(\mathbf{P}_\mu, \mathbf{P}_0) \leq \exp\left(c_2 \sum_{j=1}^{\kappa} \frac{\alpha_j^4 \beta_j^4}{\varepsilon^4}\right) - 1 \leq \exp(c_3) - 1$$

by definition of β_j . Hence, all the assumptions of Lemma A.1 are satisfied. Application of this lemma together with (E.1) implies

$$\inf_{\tilde{\mathbf{q}}} \sup_{\theta \in \Theta} \sup_{\lambda \in \mathcal{E}} \mathbf{P}(|\tilde{\mathbf{q}} - \mathbf{q}(\theta)| \geq \psi) \geq \frac{1}{4} \exp(-\beta)$$

where $\beta = \exp(c_3) - 1$, and putting this into (E.1) implies the claim assertion.

E.2 Proof of Theorem 4.2

For the proof of statement (a) we define for $\tau \in \{\pm 1\}$ hypotheses $(\theta^\tau, \lambda^\tau) \in \Theta \times \mathcal{E}$ with $\lambda^1 = \lambda^{-1} = \lambda^\circ$ for some arbitrary but fixed $\lambda^\circ \in \mathcal{E}$. Putting $\zeta = \min\{1/2, \sqrt{2}/(Ld\sqrt{\nu})\}$, for $\tau \in \{\pm 1\}$ the hypotheses concerning the solution are defined as $\theta^\tau = (\theta_j^\tau)_{j \in \mathbb{N}}$ where

$$\theta_\kappa^\tau = \frac{L}{2} (1 + \tau \zeta) \gamma_\kappa^{-1}$$

and $\theta_j^\tau = 0$ for $\tau \in \{\pm 1\}$ and $j \neq \kappa$. Then, $\theta^\tau \in \Theta$ for $\tau \in \{\pm 1\}$ since

$$\sum_{j=1}^{\infty} (\theta_j^\tau)^2 \gamma_j^2 = (\theta_\kappa^\tau)^2 \gamma_\kappa^2 \leq \frac{L^2}{4} \cdot 4 \gamma_\kappa^{-2} \gamma_\kappa^2 = L^2.$$

Denote by \mathbf{P}_τ the distribution of the tuple (X, Y) if the true parameter is $(\theta^\tau, \lambda^\tau) = (\theta^\tau, \lambda^\circ)$. Then, the Kullback–Leibler distance between \mathbf{P}_1 and \mathbf{P}_{-1} depends only on the marginal distributions \mathbf{P}_1^X and \mathbf{P}_{-1}^X , and we have by definition of ν and ζ that

$$\text{KL}(\mathbf{P}_1, \mathbf{P}_{-1}) = \frac{1}{2\varepsilon^2} \cdot (\zeta L \lambda_\kappa^\circ \gamma_\kappa^{-1})^2 \leq \frac{(\zeta L d \alpha_\kappa \gamma_\kappa^{-1})^2}{2\varepsilon^2} \leq 1.$$

Now

$$\mathfrak{q}_1 - \mathfrak{q}_{-1} = \frac{L^2}{4}(1 + \zeta)^2 \omega_\kappa^2 \gamma_\kappa^{-2} - \frac{L^2}{4}(1 - \zeta)^2 \omega_\kappa^2 \gamma_\kappa^{-2} = L^2 \zeta \omega_\kappa^2 \gamma_\kappa^{-2},$$

and (A.1) implies

$$\sup_{\theta \in \Theta} \sup_{\lambda \in \mathcal{E}} \mathbf{E}[(\tilde{\mathfrak{q}} - \mathfrak{q}(\theta))^2] \geq \frac{1}{16} L^4 \zeta^2 \omega_\kappa^4 \gamma_\kappa^{-4}$$

which implies statement (a) (again by definition of ν).

For the proof of the parametric rate ε^2 in (b) we use the same approach as in (a) but define the two hypotheses $\theta^\tau = (\theta_j^\tau)_{j \in \mathbb{N}}$, $\tau \in \{\pm 1\}$ by $\theta_1^\tau = (1 + \tau\varepsilon) \cdot \zeta$ with $\zeta = \min\{L/2, 1/(\sqrt{2}d)\}$, and $\theta_j^\tau = 0$ for $j \geq 2$. Then, $\theta^\tau \in \Theta$ since

$$\sum_{j=1}^{\infty} (\theta_j^\tau)^2 \gamma_j^2 = (1 + \tau\varepsilon)^2 \zeta^2 \leq 4 \cdot \frac{L^2}{4} = L^2$$

(recall that we assume $\varepsilon \leq 1$ throughout the paper), and the Kullback-Leibler distance between \mathbf{P}_1 and \mathbf{P}_{-1} satisfies

$$\text{KL}(\mathbf{P}_1, \mathbf{P}_{-1}) = \frac{(\lambda_1^\circ(\theta_1^1 - \theta_1^{-1}))^2}{2\varepsilon^2} \leq 2d^2 \zeta^2 \leq 1.$$

Since $\mathfrak{q}_1 - \mathfrak{q}_{-1} = 4\zeta^2\varepsilon$, the reduction scheme (A.1) implies

$$\sup_{\theta \in \Theta} \sup_{\lambda \in \mathcal{E}} \mathbf{E}[(\tilde{\mathfrak{q}} - \mathfrak{q}(\theta))^2] \geq \zeta^4 \varepsilon^2,$$

and the statement follows since $\tilde{\mathfrak{q}}$ is arbitrary.

E.3 Proof of Theorem 4.3

As in the proof of Theorem 4.2, for the proof of both parts (a) and (b) we will use the reduction scheme described in Section A.2 in the appendix where from we borrow also the notation used in the rest of the proof. In order to prove (a) define for $\tau \in \{\pm 1\}$ hypotheses $(\theta^\tau, \lambda^\tau) \in \Theta \times \mathcal{E}$ by means of

$$\begin{aligned} \theta_\kappa^\tau &= Ld^{-1}(1 + \tau\zeta)\gamma_\kappa^{-1}, & \text{and } \theta_j^\tau &= 0 \text{ for } j \neq \kappa, \\ \lambda_\kappa^\tau &= (1 - \tau\zeta)\alpha_\kappa, & \text{and } \lambda_j^\tau &= \alpha_j \text{ for } j \neq \kappa, \end{aligned}$$

where we put $\zeta = \min\{1/\sqrt{2\nu}, 1 - d^{-1}\}$. Note that the estimate $d^{-2} \leq (1 - \zeta)^2 \leq 1 \leq (1 + \zeta)^2 \leq d^2$ holds where the last inequality follows by the inequality $2d - 1 \leq d^2$ which is true for $d \geq 1$. First, $\theta^\tau \in \Theta$ for $\tau \in \{\pm 1\}$ because

$$\sum_{j=1}^{\infty} (\theta_j^\tau)^2 \gamma_j^2 = L^2 d^{-2} (1 + \tau\zeta)^2 \leq L^2.$$

Moreover $\lambda \in \mathcal{E}$, since

$$\frac{1}{d^2} \alpha_\kappa^2 \leq (1 - \zeta)^2 \alpha_\kappa^2 \leq \alpha_\kappa^2 \leq (1 + \zeta)^2 \alpha_\kappa^2 \leq d^2 \alpha_\kappa^2,$$

and $d^{-2}\alpha_j^2 \leq (\lambda_j^\circ)^2 \leq d^2\alpha_j^2$ for all $j \neq \kappa$ holds trivially. Note that $\theta^1\lambda^1 = \theta^{-1}\lambda^{-1}$ by construction, and hence the Kullback–Leibler distance between \mathbf{P}_1 and \mathbf{P}_{-1} depends only on the distance between the marginals $\mathbf{P}_1^{Y_\kappa}$ and $\mathbf{P}_{-1}^{Y_\kappa}$. Thus, by definition of ζ

$$\text{KL}(\mathbf{P}_1, \mathbf{P}_{-1}) = \text{KL}(\mathbf{P}_1^{Y_\kappa}, \mathbf{P}_{-1}^{Y_\kappa}) = \frac{1}{2\sigma^2} \cdot (2\zeta\alpha_\kappa)^2 \leq \frac{2\zeta^2\alpha_\kappa^2}{\sigma^2} \leq 1.$$

Further, it holds $\mathbf{q}_1 - \mathbf{q}_{-1} = \frac{L^2}{d^2}(1 + \zeta)^2\omega_\kappa^2\gamma_\kappa^{-2} - \frac{L^2}{d^2}(1 - \zeta)^2\omega_\kappa^2\gamma_\kappa^{-2} = \frac{4L^2}{d^2}\zeta\omega_\kappa^2\gamma_\kappa^{-2}$, and by applying (A.1) we obtain

$$\sup_{\theta \in \Theta} \sup_{\lambda \in \mathcal{E}} \mathbf{E}[(\tilde{\mathbf{q}} - \mathbf{q}(\theta))^2] \geq \frac{L^4}{d^4}\zeta^2\omega_\kappa^4\gamma_\kappa^{-4}.$$

Now (a) follows since $\sigma^2\alpha_\kappa^{-2} \asymp_\nu 1$ and $\tilde{\mathbf{q}}$ was arbitrary. For the proof of statement (b), introduce for $\tau \in \{\pm 1\}$ the hypotheses $(\theta^\tau, \lambda^\tau) \in \Theta \times \mathcal{E}$ defined by

$$\begin{aligned} \theta_1^\tau &= (1 + \tau\sigma\zeta)\frac{L}{2}, & \text{and} & & \theta_j^\tau &= 0 \text{ for } j \geq 2, \\ \lambda_1^\tau &= (1 - \tau\sigma\zeta), & \text{and} & & \lambda_j^\tau &= \alpha_j \text{ for } j \geq 2 \end{aligned}$$

where $\zeta = \min\{1/\sqrt{2}, 1 - d^{-1}\}$. Then, grant to $\sigma \leq 1$, $\theta^\tau \in \Theta$ follows from the calculation

$$\sum_{j=1}^\infty (\theta_j^\tau)^2 \gamma_j^2 \leq (1 + \tau\sigma\zeta)^2 \cdot \frac{L^2}{4} \leq L^2,$$

and the inequality $\frac{1}{d^2} \leq (1 - \sigma\zeta)^2 \leq 1 \leq (1 + \sigma\zeta)^2 \leq d^2$ shows that $\lambda \in \mathcal{E}$. By construction the Kullback–Leibler distance between \mathbf{P}_1 and \mathbf{P}_{-1} depends only on the marginal distributions of Y_1 , and hence

$$\text{KL}(\mathbf{P}_1, \mathbf{P}_{-1}) = \frac{1}{2\sigma^2}(\lambda_1^1 - \lambda_1^{-1})^2 = \frac{1}{2\sigma^2} \cdot 4\sigma^2\zeta^2 \leq 2\zeta^2 \leq 1.$$

Noting that $\mathbf{q}_1 - \mathbf{q}_{-1} = \sigma\zeta L^2$ we conclude from (A.1) that

$$\sup_{\theta \in \Theta} \sup_{\lambda \in \mathcal{E}} \mathbf{E}[(\tilde{\mathbf{q}} - \mathbf{q}(\theta))^2] \geq \frac{L^4}{16}\zeta^2\sigma^2$$

which implies the claim assertion since $\tilde{\mathbf{q}}$ was arbitrary.

APPENDIX F. PROOFS OF SECTION 5

F.1 Proof of Theorem 5.1

Consider the test statistic defined in (5.2) with

$$\tilde{C} := \max\{\sqrt{8}\delta^{-1/2}, 32d^2\delta^{-1}\}.$$

Let us first show that the type I error can be bounded by $\delta/2$. Indeed, by a direct calculation, one has

$$\mathbf{P}_0(\widehat{\Delta}^{\text{sd}} = 1) = \mathbf{P}_0(\widehat{\mathbf{q}}_{\kappa_1} \geq \tilde{C}\varphi_\varepsilon^2) \leq \frac{\mathbf{E}_0[\widehat{\mathbf{q}}_{\kappa_1}^2]}{\tilde{C}^2\varphi_\varepsilon^4} \leq \frac{2\varepsilon^4 \sum_{j=1}^{\kappa_1} \alpha_j^{-4}}{\tilde{C}^2\varphi_\varepsilon^4} \leq \delta/2.$$

where we used $\tilde{C} \geq 2\delta^{-1/2}$.

In order to bound the type II error, let $\theta \in \Theta_1(C\varphi_\varepsilon^2)$ be arbitrary. We distinguish two cases.

Case 1: $\sum_{j=1}^{\kappa_1} \theta_j^2 \geq 2d^2\tilde{C}\varphi_\varepsilon^2$. In this case we have

$$\begin{aligned} \mathbf{P}_\theta(\hat{\Delta}^{\text{sd}} = 0) &= \mathbf{P}_\theta(\hat{\mathbf{q}}_{\kappa_1} \leq \tilde{C}\varphi_\varepsilon^2) \leq \mathbf{P}_\theta\left(\hat{\mathbf{q}}_{\kappa_1} \leq \frac{1}{2d^2} \sum_{j=1}^{\kappa_1} \theta_j^2\right) \\ &\leq \mathbf{P}_\theta\left(\hat{\mathbf{q}}_{\kappa_1} - \mathbf{E}_\theta \hat{\mathbf{q}}_{\kappa_1} \leq -\frac{1}{2d^2} \sum_{j=1}^{\kappa_1} \theta_j^2\right) \end{aligned}$$

where we have used that $\mathbf{E}_\theta \hat{\mathbf{q}}_{\kappa_1} \geq \sum_{j=1}^{\kappa_1} \theta_j^2/d^2$. Thus,

$$\begin{aligned} \mathbf{P}_\theta(\hat{\Delta}^{\text{sd}} = 0) &\leq \frac{4d^4 \mathbf{E}_\theta[(\hat{\mathbf{q}}_{\kappa_1} - \mathbf{E}_\theta \hat{\mathbf{q}}_{\kappa_1})^2]}{(\sum_{j=1}^{\kappa_1} \theta_j^2)^2} \\ &\leq 4d^4 \left\{ \frac{2\varepsilon^4 \sum_{j=1}^{\kappa_1} \alpha_j^{-4}}{(\sum_{j=1}^{\kappa_1} \theta_j^2)^2} + \frac{4\varepsilon^2 \sum_{j=1}^{\kappa_1} \alpha_j^{-4} (\lambda_j \theta_j)^2}{(\sum_{j=1}^{\kappa_1} \theta_j^2)^2} \right\} \\ &\leq 4d^4 \left\{ \frac{2}{4d^4 \tilde{C}^2} + \frac{4\varepsilon^2 \alpha_{\kappa_1}^{-2} \sum_{j=1}^{\kappa_1} \theta_j^2}{(\sum_{j=1}^{\kappa_1} \theta_j^2)^2} \right\} \\ &\leq \frac{2}{\tilde{C}^2} + \frac{16d^4 \varepsilon^2 \alpha_{\kappa_1}^{-2}}{2d^2 \tilde{C} \varphi_\varepsilon^2} \\ &\leq \frac{2}{\tilde{C}^2} + \frac{8d^2}{\tilde{C}} \\ &\leq \delta/2 \end{aligned}$$

where the last estimate is due to $\tilde{C} \geq \max\{\sqrt{8}\delta^{-1/2}, 32d^2\delta^{-1}\}$. Since $\theta \in \Theta_1$ was arbitrary, this shows that the type II error can be bounded from above by $\delta/2$ in this case.

Case 2: $\sum_{j=1}^{\kappa_1} \theta_j^2 \leq 2d^2\tilde{C}\varphi_\varepsilon^2$. First note that, by definition of Θ_1 ,

$$\begin{aligned} \mathbf{E}_\theta \hat{\mathbf{q}}_{\kappa_1} &\geq \frac{1}{d^2} \sum_{j=1}^{\kappa_1} \theta_j^2 = \frac{1}{d^2} \left\{ \sum_{j=1}^{\infty} \theta_j^2 - \sum_{j>\kappa_1} \theta_j^2 \right\} \\ &\geq \frac{1}{d^2} \left(C^2 \varphi_\varepsilon^2 - \sum_{j>\kappa_1} \frac{\gamma_j^2}{\gamma_j^2} \theta_j^2 \right) \\ &\geq \left(\frac{C^2}{d^2} - \frac{L^2 \sqrt{\nu}}{d^2} \right) \varphi_\varepsilon^2. \end{aligned}$$

Now, since $\varepsilon^2 \sum_{j=1}^{\kappa_1} \alpha_j^{-2} \theta_j^2 \leq \varepsilon^2 \alpha_{\kappa_1}^{-2} \sum_{j=1}^{\kappa_1} \theta_j^2 \leq 2d^2 \tilde{C} \varphi_\varepsilon^4$ in Case 2, we obtain, choosing C such that $C^2/d^2 - L^2\sqrt{\nu}/d^2 - \tilde{C} > 0$,

$$\begin{aligned} \mathbf{P}_\theta(\hat{\Delta}^{\text{sd}} = 0) &= \mathbf{P}_\theta(\hat{\mathbf{q}}_{\kappa_1} - \mathbf{E}_\theta \hat{\mathbf{q}}_{\kappa_1} \leq \tilde{C}\varphi_\varepsilon^2 - \mathbf{E}_\theta \hat{\mathbf{q}}_{\kappa_1}) \\ &\leq \mathbf{P}_\theta(\mathbf{E}_\theta \hat{\mathbf{q}}_{\kappa_1} - \hat{\mathbf{q}}_{\kappa_1} \geq (C^2/d^2 - L^2\sqrt{\nu}/d^2 - \tilde{C})\varphi_\varepsilon^2) \end{aligned}$$

$$\begin{aligned}
 &\leq \frac{\text{Var}(\widehat{\mathbf{q}}_{\kappa_1})}{\left(\frac{C^2}{d^2} - \widetilde{C} - \frac{L^2\sqrt{\nu}}{d^2}\right)^2} \varphi_\varepsilon^4 \\
 &\leq \frac{2\varepsilon^4 \sum_{j=1}^{\kappa_1} \alpha_j^{-4}}{\left(\frac{C^2}{d^2} - \widetilde{C} - \frac{L^2\sqrt{\nu}}{d^2}\right)^2} \varphi_\varepsilon^4 + \frac{4d^2\varepsilon^2 \sum_{j=1}^{\kappa_1} \alpha_j^{-2} \theta_j^2}{\left(\frac{C^2}{d^2} - \widetilde{C} - \frac{L^2\sqrt{\nu}}{d^2}\right)^2} \varphi_\varepsilon^4 \\
 &\leq \frac{2}{\left(\frac{C^2}{d^2} - \widetilde{C} - \frac{L^2\sqrt{\nu}}{d^2}\right)^2} + \frac{8d^4\widetilde{C}}{\left(\frac{C^2}{d^2} - \widetilde{C} - \frac{L^2\sqrt{\nu}}{d^2}\right)^2}
 \end{aligned}$$

and the last expression is bounded from above by $\delta/2$ for C sufficiently large. Thus, the type II error is bounded by $\delta/2$ also in Case 2 and the statement of the proposition follows.

F.2 Proof of Theorem 5.2

In order to prove the theorem, we will use statement (i) from Lemma A.2. For any $\tau \in \{\pm 1\}^{\kappa_1}$ define θ^τ by

$$\theta_i^\tau = \tau_i \varepsilon c \cdot \frac{\alpha_i^{-2}}{\left(\sum_{j=1}^{\kappa_1} \alpha_j^{-4}\right)^{1/4}}, \quad i \in \llbracket 1, \kappa_1 \rrbracket,$$

and $\theta_i^\tau = 0$ for $i > \kappa_1$. Then, in analogy to the proof of Theorem 4.1, it can be shown that $\theta^\tau \in \Theta$ provided that $c^2 \leq L^2\nu^{-1/2}$. Moreover, for all $\tau \in \{\pm 1\}^{\kappa_1}$,

$$\mathbf{q}(\theta^\tau) = c^2\varepsilon^2 \sqrt{\sum_{j=1}^{\kappa_1} \alpha_j^{-4}} = c^2\varphi_\varepsilon^2,$$

and hence the law of κ_1 independent Rademacher random variables induces a probability distribution μ on the set $\Theta_1(c\varphi_\varepsilon)$. Finally, again in analogy to the proof of Theorem 4.1, it holds

$$\chi^2(\mathbf{P}_0, \mathbf{P}_\mu) \leq \exp(c_2c^2) - 1$$

for some fixed numerical constant $c_2 = c_2(d) > 0$. Now, taking c sufficiently small implies $\chi^2(\mathbf{P}_0, \mathbf{P}_\mu) \leq \delta^2$, and applying Lemma A.2 yields the claim assertion.

F.3 Proof of Theorem 5.3

We consider the test statistic defined in (5.3) where the conditions on \widetilde{C} will be stated in the sequel. We start by bounding the type I error from above by $\delta/2$:

$$\mathbf{P}_0(\widehat{\Delta}^{\text{gof}} = 1) = \mathbf{P}_0(\widehat{\mathbf{q}}_{\kappa_2} \geq \widetilde{C}\varphi_{\varepsilon,\sigma}^2) \leq \frac{\mathbf{E}_0[\widehat{\mathbf{q}}_{\kappa_2}^2]}{\widetilde{C}^2\varphi_{\varepsilon,\sigma}^4}.$$

Now,

$$\begin{aligned}
 \mathbf{E}_0[\widehat{\mathbf{q}}_{\kappa_2}^2] &\leq 672d^4\varepsilon^4 \sum_{j=1}^{\kappa_2} \alpha_j^{-4} + 2688d^4L^4\sigma^4 \max_{j \in \llbracket 1, \kappa_2 \rrbracket} \alpha_j^{-4} \gamma_j^{-4} \\
 &\leq (672d^4 + 2688d^4L^4)\varphi_{\varepsilon,\sigma}^4,
 \end{aligned}$$

and hence $\mathbf{P}_0(\widehat{\Delta}^{\text{gof}} = 1) \leq \delta/2$ provided that $\widetilde{C}^2 \geq 2(672d^4 + 2688d^4L^4)\delta^{-1}$.

Now, we consider the type II error. In order to bound it from above by $\delta/2$, let $\theta \in \Theta_1(C\varphi_{\varepsilon,\sigma})$ be arbitrary. It holds

$$\mathbf{P}_\theta(\widehat{\Delta}^{\text{gof}} = 0) = \mathbf{P}_\theta(\widehat{\mathbf{q}}_{\kappa_2} \leq \widetilde{C}\varphi_{\varepsilon,\sigma}^2) = \mathbf{P}_\theta(\widehat{\mathbf{q}}_{\kappa_2} - \sum_{j=1}^{\kappa_2}(\theta_j - \theta_j^\circ)^2 \leq \widetilde{C}\varphi_{\varepsilon,\sigma}^2 - \sum_{j=1}^{\kappa_2}(\theta_j - \theta_j^\circ)^2),$$

and as in the proof of Theorem 5.1 we consider two cases.

Case 1: $\sum_{j=1}^{\kappa_2}(\theta_j - \theta_j^\circ)^2 \geq 2\widetilde{C}\varphi_{\varepsilon,\sigma}^2$. Then $\widetilde{C}\varphi_{\varepsilon,\sigma}^2 \leq \sum_{j=1}^{\kappa_2}(\theta_j - \theta_j^\circ)^2/2$, and thus

$$\begin{aligned} \mathbf{P}_\theta(\widehat{\Delta}^{\text{gof}} = 0) &\leq \mathbf{P}_\theta\left(\widehat{\mathbf{q}}_{\kappa_2} - \sum_{j=1}^{\kappa_2}(\theta_j - \theta_j^\circ)^2 \leq -\sum_{j=1}^{\kappa_2}(\theta_j - \theta_j^\circ)^2/2\right) \\ &= \mathbf{P}_\theta\left(-\widehat{\mathbf{q}}_{\kappa_2} + \sum_{j=1}^{\kappa_2}(\theta_j - \theta_j^\circ)^2 \geq \sum_{j=1}^{\kappa_2}(\theta_j - \theta_j^\circ)^2/2\right) \\ &\leq \frac{4\mathbf{E}_\theta[(\widehat{\mathbf{q}}_{\kappa_2} - \sum_{j=1}^{\kappa_2}(\theta_j - \theta_j^\circ)^2)^2]}{(\sum_{j=1}^{\kappa_2}(\theta_j - \theta_j^\circ)^2)^2}. \end{aligned}$$

Now, similarly as in the proof of Theorem 3.1,

$$\mathbf{E}_\theta[(\widehat{\mathbf{q}}_{\kappa_2} - \sum_{j=1}^{\kappa_2}(\theta_j - \theta_j^\circ)^2)^2] \leq 3 \sum_{i=1}^3 \mathbf{E}\mathcal{T}_{\kappa_2 i}^2$$

where $\mathcal{T}_{\kappa_2 1}$, $\mathcal{T}_{\kappa_2 2}$, and $\mathcal{T}_{\kappa_2 3}$ are defined as in the proof of Theorem 3.1. Following line by line the derivation of the upper bounds for the three terms on the right-hand side of the last display from Appendix C, we obtain

$$\mathbf{E}_\theta[(\widehat{\mathbf{q}}_{\kappa_2} - \sum_{j=1}^{\kappa_2}(\theta_j - \theta_j^\circ)^2)^2] \leq C(d, L)\varphi_{\varepsilon,\sigma}^4 + C(d, L)\varphi_{\varepsilon,\sigma}^2 \sum_{j=1}^{\kappa_2}(\theta_j - \theta_j^\circ)^2.$$

Hence

$$\begin{aligned} \mathbf{P}_\theta(\widehat{\Delta}^{\text{gof}} = 0) &\leq \frac{C(d, L)\varphi_{\varepsilon,\sigma}^4}{\widetilde{C}^2\varphi_{\varepsilon,\sigma}^4} + \frac{C(d, L)\varphi_{\varepsilon,\sigma}^2}{\sum_{j=1}^{\kappa_2}(\theta_j - \theta_j^\circ)^2} \\ &\leq \frac{C(d, L)}{\widetilde{C}^2} + \frac{C(d, L)}{\widetilde{C}}, \end{aligned}$$

and the last expression is smaller than $\delta/2$ for \widetilde{C} sufficiently large¹.

Case 2: $\sum_{j=1}^{\kappa_2}(\theta_j - \theta_j^\circ)^2 \leq 2\widetilde{C}\varphi_{\varepsilon,\sigma}^2$. First note that $\theta \in \Theta_1(C\varphi_{\varepsilon,\sigma})$ implies

$$\sum_{j=1}^{\kappa_2}(\theta_j - \theta_j^\circ)^2 = \sum_{j=1}^{\infty}(\theta_j - \theta_j^\circ)^2 - \sum_{j>\kappa_2}(\theta_j - \theta_j^\circ)^2 \geq (C^2 - 2L^2\sqrt{\nu})\varphi_{\varepsilon,\sigma}^2.$$

¹In order to make a lower bound on \widetilde{C} explicit, it would be necessary to make the constants in statement (ii) of Proposition C.1 explicit, and we do not address this issue here.

Thus, for C sufficiently large²,

$$\begin{aligned}
 \mathbf{P}_\theta(\widehat{\Delta}^{\text{gof}} = 0) &\leq \mathbf{P}_\theta\left(\widehat{\mathbf{q}}_{\kappa_2} - \sum_{j=1}^{\kappa_2}(\theta_j - \theta_j^\circ)^2 \leq \widetilde{C}\varphi_{\varepsilon,\sigma}^2 - \sum_{j=1}^{\kappa_2}(\theta_j - \theta_j^\circ)^2\right) \\
 &\leq \mathbf{P}_\theta\left(\widehat{\mathbf{q}}_{\kappa_2} - \sum_{j=1}^{\kappa_2}(\theta_j - \theta_j^\circ)^2 \leq (\widetilde{C} - C^2 + 2L^2\sqrt{\nu})\varphi_{\varepsilon,\sigma}^2\right) \\
 &= \mathbf{P}_\theta\left(-\widehat{\mathbf{q}}_{\kappa_2} + \sum_{j=1}^{\kappa_2}(\theta_j - \theta_j^\circ)^2 \geq (C^2 - \widetilde{C} - 2L^2\sqrt{\nu})\varphi_{\varepsilon,\sigma}^2\right) \\
 &\leq \frac{\mathbf{E}_\theta[(\widehat{\mathbf{q}}_{\kappa_2} - \sum_{j=1}^{\kappa_2}(\theta_j - \theta_j^\circ)^2)^2]}{(C^2 - \widetilde{C} - 2L^2\sqrt{\nu})^2\varphi_{\varepsilon,\sigma}^4}.
 \end{aligned}$$

Now, as in the first case,

$$\mathbf{E}_\theta[(\widehat{\mathbf{q}}_{\kappa_2} - \sum_{j=1}^{\kappa_2}(\theta_j - \theta_j^\circ)^2)^2] \leq C(d, L)\varphi_{\varepsilon,\sigma}^4 + C(d, L)\varphi_{\varepsilon,\sigma}^2 \sum_{j=1}^{\kappa_2}(\theta_j - \theta_j^\circ)^2 \leq C(d, L)\varphi_{\varepsilon,\sigma}^4.$$

Hence,

$$\mathbf{P}_\theta(\widehat{\Delta}^{\text{gof}} = 0) \leq \frac{C(d, L)\varphi_{\varepsilon,\sigma}^4}{(C^2 - \widetilde{C} - 2L^2\sqrt{\nu})^2\varphi_{\varepsilon,\sigma}^4}$$

and $\mathbf{P}_\theta(\widehat{\Delta}^{\text{gof}} = 0) \leq \delta/2$ provided that C is sufficiently large³.

F.4 Proof of Theorem 5.5

The case that $\varphi_{\varepsilon,\sigma}^2 = \varepsilon^2 \sqrt{\sum_{j=1}^{\kappa_2} \alpha_j^{-4}}$ is dealt with in analogy to the proof of Theorem 5.2, and thus omitted (the additional assumption $\varphi_{\varepsilon,\sigma}^4 \asymp_\nu \gamma_{\kappa_2}^{-4}$ is only exploited in this case). Thus, we consider the case $\varphi_{\varepsilon,\sigma}^2 = \sigma^2 \max_{j \in \llbracket 1, \kappa_2 \rrbracket} \alpha_j^{-2} \gamma_j^{-2}$, and put $\kappa = \operatorname{argmax}_{j \in \llbracket 1, \kappa_2 \rrbracket} \alpha_j^{-2} \gamma_j^{-2}$. We apply statement (ii) of Lemma A.2 to the testing problem

$$\mathcal{H}_0: \theta = \theta^\circ, \lambda = \lambda^\circ \quad \text{against} \quad \mathcal{H}_1: \theta = \theta^1, \lambda = \lambda^1$$

where $\theta_j^1 = \theta_j^\circ$ for $j \neq \kappa$, $\theta_\kappa^1 = \frac{1 - \widetilde{c}\sigma\alpha_\kappa^{-1}\gamma_\kappa^{-1}}{1 + \widetilde{c}\sigma\alpha_\kappa^{-1}\gamma_\kappa^{-1}} \cdot \theta_\kappa^\circ$, $\lambda_j^\circ = \lambda_j^1 = \alpha_j$ for $j \neq \kappa$, $\lambda_\kappa^\circ = (1 - \widetilde{c}\sigma\alpha_\kappa^{-1}\gamma_\kappa^{-1})\alpha_\kappa$, and $\lambda_\kappa^1 = (1 + \widetilde{c}\sigma\alpha_\kappa^{-1}\gamma_\kappa^{-1})\alpha_\kappa$. First, it is easily checked that $\theta^1 \in \Theta$ and $\lambda^\circ, \lambda^1 \in \mathcal{E}$ for \widetilde{c} sufficiently small. Further, since $\sigma\alpha_\kappa^{-1}\gamma_\kappa^{-1} \leq 1$ by definition of $\varphi_{\varepsilon,\sigma}$,

$$\|\theta^\circ - \theta^1\|_2^2 = (\theta_\kappa^\circ - \theta_\kappa^1)^2 = \frac{4\widetilde{c}^2\sigma^2\alpha_\kappa^{-2}\gamma_\kappa^{-2}}{1 + \widetilde{c}\sigma\alpha_\kappa^{-1}\gamma_\kappa^{-1}} \cdot (\theta_\kappa^\circ)^2 \geq \frac{4\widetilde{c}^2}{1 + \widetilde{c}} \cdot (\theta_\kappa^\circ)^2 \cdot \varphi_{\varepsilon,\sigma}^2 =: c^2\varphi_{\varepsilon,\sigma}^2,$$

and $c \rightarrow 0$ if and only if $\widetilde{c} \rightarrow 0$, showing that $\theta^1 \in \Theta_1(c\varphi_{\varepsilon,\sigma})$ for \widetilde{c} sufficiently small. Thus, it remains to show that the Kullback-Leibler distance between the two hypotheses can be made arbitrary small by choosing the

²Again, we are not able to give explicit bounds on C due to the fact that the constant in statement (ii) of Proposition C.1 is not made explicit.

³See Footnote 2.

parameter \tilde{c} sufficiently small. By construction, $\text{KL}(\mathbf{P}_0^{X,Y}, \mathbf{P}_1^{X,Y}) = \text{KL}(\mathbf{P}_0^{Y_\kappa}, \mathbf{P}_1^{Y_\kappa})$, and hence

$$\text{KL}(\mathbf{P}_0^{X,Y}, \mathbf{P}_1^{X,Y}) = \frac{2}{\sigma^2} \tilde{c}^2 \sigma^2 \gamma_\kappa^{-2} \alpha_\kappa^{-2} \alpha_\kappa^2 \leq 2\tilde{c}^2,$$

and $2\tilde{c}^2 \leq 2\delta^2 \Leftrightarrow \tilde{c} \leq \delta$ implies the claim assertion grant to statement (ii) of Lemma A.2 with $\mu = \delta_{(\theta^1, \lambda^1)}$.

REFERENCES

- [1] P.J. Bickel and Y. Ritov, Estimating integrated squared density derivatives: sharp best order of convergence estimates. *Sankhyā Ser. A* **50** (1988) 381–393.
- [2] N. Bissantz, G. Claeskens, H. Holzmann and A. Munk, Testing for lack of fit in inverse regression—with applications to biophotonic imaging. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **71** (2009) 25–48.
- [3] C. Butucea, Goodness-of-fit testing and quadratic functional estimation from indirect observations. *Ann. Stat.* **35** (2007) 1907–1930.
- [4] C. Butucea, C. Matias and C. Pouet, Adaptive goodness-of-fit testing from indirect observations. *Ann. Inst. Henri Poincaré Probab. Stat.* **45** (2009) 352–372.
- [5] C. Butucea and K. Meziani, Quadratic functional estimation in inverse problems. *Stat. Methodol.* **8** (2011) 31–41.
- [6] T.T. Cai and M.G. Low, Nonquadratic estimators of a quadratic functional. *Ann. Stat.* **33** (2005) 2930–2956.
- [7] T.T. Cai and M.G. Low, Optimal adaptive estimation of a quadratic functional. *Ann. Stat.* **34** (2006) 2298–2325.
- [8] L. Cavalier, Inverse problems in statistics. In *Inverse problems and high-dimensional estimation*, Vol. 203 of *Lect. Notes Stat. Proc.* Springer, Heidelberg (2011) 3–96.
- [9] L. Cavalier and N.W. Hengartner, Adaptive estimation for inverse problems with noisy operators. *Inverse Probl.* **21** (2005) 1345–1361.
- [10] C. Chesneau, On adaptive wavelet estimation of a quadratic functional from a deconvolution problem. *Ann. Inst. Stat. Math.* **63** (2011) 405–429.
- [11] O. Collier, L. Comminges and A.B. Tsybakov, Minimax estimation of linear and quadratic functionals on sparsity classes. *Ann. Stat.* **45** (2017) 923–958.
- [12] F. Comte and C. Lacour, Data-driven density estimation in the presence of additive noise with unknown distribution. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **73** (2011) 601–627.
- [13] D.L. Donoho, Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition. *Appl. Comput. Harmon. Anal.* **2** (1995) 101–126.
- [14] D.L. Donoho and M. Nussbaum, Minimax quadratic estimation of a quadratic functional. *J. Complex.* **6** (1990) 290–323.
- [15] S. Efromovich and V. Koltchinskii, On inverse problems with unknown operators. *IEEE Trans. Inf. Theory* **47** (2001) 2876–2894.
- [16] M.S. Ermakov, Minimax detection of a signal in Gaussian white noise. *Teor. Veroyatnost. i Primenen.* **35** (1990) 704–715.
- [17] J. Fan, On the estimation of quadratic functionals. *Ann. Stat.* **19** (1991) 1273–1294.
- [18] J. Fan and I. Gijbels, Minimax estimation of a bounded squared mean. *Stat. Probab. Lett.* **13** (1992) 383–390.
- [19] G. Gayraud and K. Tribouley, Wavelet methods to estimate an integrated quadratic functional: adaptivity and asymptotic law. *Stat. Probab. Lett.* **44** (1999) 109–122.
- [20] E. Giné and R. Nickl, A simple adaptive estimator of the integrated square of a density. *Bernoulli* **14** (2008) 47–61.
- [21] M. Hoffmann and M. Reiss, Nonlinear estimation for linear inverse problems with error in the operator. *Ann. Stat.* **36** (2008) 310–336.
- [22] Y.I. Ingster and I.A. Suslina, Nonparametric goodness-of-fit testing under Gaussian models. In Vol. 169 of *Lecture Notes in Statistics*. Springer-Verlag, New York (2003).
- [23] Y.I. Ingster, Asymptotically minimax hypothesis testing for nonparametric alternatives (i–iii). *Math. Methods Stat.* **2** (1993) 85–114, 171–189, 249–268.
- [24] Y.I. Ingster, T. Sapatinas and I.A. Suslina, Minimax signal detection in ill-posed inverse problems. *Ann. Stat.* **40** (2012) 1524–1549.
- [25] J. Johannes, Deconvolution with unknown error distribution. *Ann. Stat.* **37** (2009) 2301–2323.
- [26] J. Johannes and M. Schwarz, Adaptive Gaussian inverse regression with partially unknown operator. *Commun. Stat. Theory Methods* **42** (2013) 1343–1362.
- [27] I. Johnstone, Thresholding for weighted χ^2 . *Stat. Sinica* **11** (2001) 691–704.
- [28] I.M. Johnstone, Chi-square oracle inequalities. In State of the art in probability and statistics (Leiden, 1999). In Vol. 36 of *IMS Lecture Notes Monogr. Ser.* Inst. Math. Stat., Beachwood, OH (2001) 399–418.
- [29] J. Klemelä, Sharp adaptive estimation of quadratic functionals. *Probab. Theory Related Fields* **134** (2006) 539–564.
- [30] C. Lacour and T.M. Pham Ngoc, Goodness-of-fit test for noisy directional data. *Bernoulli* **20** (2014) 2131–2168.
- [31] B. Laurent, J.-M. Loubes and C. Marteau, Testing inverse problems: a direct or an indirect problem? *J. Stat. Plann. Infer.* **141** (2011) 1849–1861.
- [32] B. Laurent and P. Massart, Adaptive estimation of a quadratic functional by model selection. *Ann. Stat.* **28** (2000) 1302–1338.

- [33] B. Laurent, Adaptive estimation of a quadratic functional of a density by model selection. *ESAIM: PS* **9** (2005) 1–18.
- [34] C. Marteau and T. Sapatinas, A unified treatment for non-asymptotic and asymptotic approaches to minimax signal detection. *Stat. Surv.* **9** (2015) 253–297.
- [35] C. Marteau and T. Sapatinas, Minimax goodness-of-fit testing in ill-posed inverse problems with partially unknown operators. *Ann. Inst. Henri Poincaré Probab. Stat.* **53** (2017) 1675–1718.
- [36] M.H. Neumann, On the effect of estimating the error density in nonparametric deconvolution. *J. Nonparametr. Stat.* **7** (1997) 307–330.
- [37] V. Rivoirard and K. Tribouley, The maxiset point of view for estimating integrated quadratic functionals. *Stat. Sinica* **18** (2008) 255–279.
- [38] A.B. Tsybakov, Introduction to nonparametric estimation. In *Springer Series in Statistics*. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats. Springer, New York (2009).
- [39] A.B. Tsybakov, Aggregation and minimax optimality in high-dimensional estimation. In *Proc. of the International Congress of Mathematicians—Seoul 2014. Vol. IV*. Kyung Moon Sa, Seoul (2014) 225–246.