Minimax Signal Detection under Weak Noise Assumptions

C. Marteau[1*] and Th. Sapatinas[2**]

[1]*Univ. Lyon, Université Claude Bernard Lyon 1,
CNRS UMR 5208, Institut Camille Jordan, France*
[2]*Dept. Math. and Statist. Univ. of Cyprus, Nicosia, Cyprus*
Received January 13, 2017; in final form, September 19, 2017

Abstract—We consider minimax signal detection in the sequence model. Working with certain ellipsoids in the space of square-summable sequences of real numbers, with a ball of positive radius removed, we obtain upper and lower bounds for the minimax separation radius in the non-asymptotic framework, i.e., for a fixed value of the involved noise level. We use very weak assumptions on the noise (i.e., fourth moments are assumed to be uniformly bounded). In particular, we do not use any kind of Gaussian distribution or independence assumption on the noise. It is shown that the established minimax separation rates are not faster than the ones obtained in the classical sequence model (i.e., independent standard Gaussian noise) but, surprisingly, are of the same order as the minimax estimation rates in the classical setting. Under an additional condition on the noise, the classical minimax separation rates are also retrieved in benchmark well-posed and ill-posed inverse problems.

Keywords: ellipsoids, ill-posed inverse problems, minimax signal detection, well-posed inverse problems.

AMS 2010 Subject Classification: 62G05, 62K20.

DOI: 10.3103/S1066530717040032

1. INTRODUCTION

We consider the following sequence model (SM),

$$y_k = b_k \theta_k + \varepsilon \, \xi_k, \quad k \in \mathcal{N}, \tag{1.1}$$

where $\mathcal{N}$ can be either $\mathbb{N} = \{1, 2, \dots\}$ or $\mathbb{N}_n = \{1, \dots, n\}$ for some $n \geq 1$, $b = (b_k)_{k \in \mathcal{N}}$ is a known positive sequence, $\theta = (\theta_k)_{k \in \mathcal{N}} \in l^2(\mathcal{N})$ is the unknown signal of interest, $\xi = (\xi_k)_{k \in \mathcal{N}}$ is a sequence of random variables (the noise), and $\varepsilon > 0$ is a known parameter (the noise level). The observations are given by the sequence $y = \{y_k\}_{k \in \mathcal{N}}$ from the SM (1.1) and their joint law is denoted by $\mathbb{P}_{\theta,\xi}$. Here, $l^2(\mathcal{N})$ denotes the space of squared-summable sequence of real numbers, i.e.,

$$l^2(\mathcal{N}) = \left\{ \theta \in \mathbb{R}^{\mathcal{N}} : \|\theta\|^2 := \sum_{k \in \mathcal{N}} \theta_j^2 < +\infty \right\}.$$

Let $\mathcal{C} > 0$ be a known fixed constant. Concerning the noise, we will assume that $\xi \in \Xi$, where

$$\Xi := \Xi(\mathcal{C}) = \left\{ \xi : \mathbb{E}[\xi_k] = 0, \ \mathbb{E}[\xi_k^2] = 1 \ \forall k \in \mathcal{N} \ \text{and} \ \sup_{k \in \mathcal{N}} \mathbb{E}[\xi_k^4] \leq \mathcal{C} < +\infty \right\}. \tag{1.2}$$

The SM (1.1) arises in many well-known situations. Consider for instance the stochastic differential equation

$$dZ_\varepsilon(t) = Af(t) + \epsilon dU(t), \quad t \in [0, 1],$$

*E-mail: `marteau@math.univ-lyon1.fr`
**E-mail: `fanis@ucy.ac.cy`

where A is a known bounded linear operator acting on $L^2([0,1])$, $f(\cdot) \in L^2([0,1])$ is the unknown response function that one wants to detect or estimate, $U(\cdot)$ is a given stochastic process on $[0,1]$ and $\varepsilon > 0$ is a known parameter (the noise level). For the sake of simplicity, we only consider the case when A is injective (meaning that A has a trivial nullspace).

- Let $U(\cdot) = W(\cdot)$ be the standard Wiener process. Then, if A is the identity operator, we can retrieve the SM (1.1) in the Fourier domain with $b_k = 1$ for all $k \in \mathcal{N} = \mathbb{N}$ and the ξ_k, $k \in \mathcal{N}$, are independent standard Gaussian random variables (direct problem). If A is a self-adjoint operator with an eigen-decomposition, we can retrieve the SM (1.1) where $b_k > b_0$ for some $b_0 > 0$ for all $k \in \mathcal{N} = \mathbb{N}$ and the ξ_k, $k \in \mathcal{N}$, are independent standard Gaussian random variables (well-posed inverse problems). If A is a compact operator, we can retrieve the SM (1.1) where $b_k > 0$ for all $k \in \mathcal{N}$ (since A is injective) with $b_k \to 0$ as $k \to +\infty$ and the ξ_k, $k \in \mathcal{N}$, are independent standard Gaussian random variables (ill-posed inverse problems). For more details regarding all these models, we refer to, e.g., [5].

- Let $U(\cdot) = W_{-\gamma}(\cdot)$, $\gamma \in]0, 1/2[$, be the truncated fractional Brownian motion and let A be the identity operator. Then we can retrieve the SM (1.1) in the spline domain with $b_k = (\pi k)^{-2\gamma}(1 + o(1))$ as $k \to +\infty$ and the $\xi_k, k \in \mathcal{N}$, are (non-independent) standard Gaussian random variables. For more details, we refer to, e.g., [10], [4].

The non-parametric inverse regression problem also provides observations of the form (1.1). Indeed, consider the model

$$Z_i = Af\left(\frac{i}{n}\right) + \frac{1}{\sqrt{n}}\,\eta_i, \quad i \in \{1, \ldots, n\},$$

where A is a known (injective) bounded linear operator acting on $L^2([0,1])$, $f(\cdot) \in L^2([0,1])$ is the unknown response function that one wants to detect or estimate, and η_i, $i \in \mathbb{N}_n$, is a sequence of independent and identically distributed random variables with zero mean, variance one and finite fourth moment. Given any appropriate bases (or, even, a tight frame, see, e.g., [13], p. 126), we can retrieve the SM (1.1) with $b_k = 1$ for all $k \in \mathcal{N}_n$ when A is the identity operator, see, e.g., [16], Chapter 1. When A is a compact operator, we can retrieve an approximation of the SM (1.1) where $(b_k)_{k \in \mathbb{N}_n}$ is a fixed sequence that depends on A, see, e.g., [2].

Minimax signal detection has been considered in the literature over the last two decades. We refer to, e.g., [1], [12], [8], [11], [7], [6], [14]. All these contributions consider the classical Gaussian sequence model (1.1), i.e., where the ξ_k, $k \in \mathcal{N}$, are independent standard Gaussian random variables. We refer to [15] for a survey on available results and a discussion on the link between asymptotic (the noise level is assumed to tend to zero) and non-asymptotic (the noise level is assumed to be fixed) approaches to minimax signal detection. The aim of this work is to obtain upper and lower bounds on the minimax separation radius in the non-asymptotic framework, for the general model (1.1) under weak assumptions on the noise, i.e., when $\xi \in \Xi$, where the set Ξ has been introduced in (1.2). In particular, we do not use any kind of Gaussian distribution or independence assumption on the noise. We prove that the minimax separation rates are not faster than the ones obtained in the classical sequence model (see, e.g., [1], [12], [8], [11], [7], [6], [14]) but, surprisingly, are of the same order as the minimax estimation rates in the classical setting. Moreover, under additional conditions on the noise, we show that the classical minimax separation rates can be retrieved in benchmark well-posed and ill-posed inverse problems.

Throughout the paper, we use the following notation. Given two sequences $(c_k)_{k \in \mathcal{N}}$ and $(d_k)_{k \in \mathcal{N}}$ of real numbers, $c_k \sim d_k$ means that there exist $0 < \kappa_0 \leq \kappa_1 < \infty$ such that $\kappa_0 \leq c_k/d_k \leq \kappa_1$ for all $k \in \mathcal{N}$, while $c_k \lesssim d_k$ (resp. $c_k \gtrsim d_k$) means $c_k \leq c_0\, d_k$ (resp. $c_k \geq c_0\, d_k$) for some $c_0 > 0$ for all $k \in \mathcal{N}$. Also, $x \wedge y := \min(x, y)$ for all $x, y \in \mathbb{R}$.

2. MINIMAX SIGNAL DETECTION

Given observations from the SM (1.1), we consider the signal detection problem, i.e., our aim is to test

$$H_0 \colon \theta = 0 \quad \text{versus} \quad H_1 \colon \theta \in \Theta_a(r_\varepsilon). \tag{2.1}$$

Given a non-decreasing sequence $a = (a_k)_{k \in \mathcal{N}}$ of positive real numbers with $a_k \to +\infty$ as $k \to +\infty$ when $\mathcal{N} = \mathbb{N}$ and a radius $r_\varepsilon > 0$, the set $\Theta_a(r_\varepsilon)$ is defined as

$$\Theta_a(r_\varepsilon) = \big\{ \theta \in \mathcal{E}_a, \ \|\theta\| \geq r_\varepsilon \big\}, \tag{2.2}$$

where

$$\mathcal{E}_a = \Big\{ \theta \in l^2(\mathcal{N}), \ \sum_{k \in \mathcal{N}} a_k^2 \theta_k^2 \leq 1 \Big\}.$$

The set $\mathcal{E}_a$ can be viewed as a condition on the decay of θ. The cases where the sequence a increases very fast correspond to the signal θ with small coefficients. In such a case, the corresponding signal can be considered as being 'smooth'. The sequence a being fixed, the main issue for the minimax signal detection problem (2.1)–(2.2) is then to characterize the values of the radius $r_\varepsilon > 0$ for which both hypotheses H_0 (called the null hypothesis) and H_1 (called the alternative hypothesis) are 'separable'.

In the following, a (non-randomized) test $\Psi := \Psi(y)$ will be defined as a measurable function of the observation $y = (y_k)_{k \in \mathcal{N}}$ from the SM (1.1) having values in the set $\{0, 1\}$. By convention, H_0 is rejected if $\Psi = 1$ and H_0 is not rejected if $\Psi = 0$. Then, given a test Ψ, we can investigate

- the type I (first kind) error probability defined as

$$\sup_{\xi \in \Xi} \mathbb{P}_{0,\xi}(\Psi = 1), \tag{2.3}$$

  which measures the worst probability of rejecting H_0 when H_0 is true (i.e., $\theta = 0$, $\xi \in \Xi$, where Ξ is defined in (1.2)); it is often constrained as being bounded by a prescribed level $\alpha \in]0, 1[$, and

- the type II (second kind) error probability defined as

$$\sup_{\substack{\theta \in \Theta_a(r_\varepsilon) \\ \xi \in \Xi}} \mathbb{P}_{\theta,\xi}(\Psi = 0), \tag{2.4}$$

  which measures the worst possible probability of not rejecting H_0 when H_0 is not true (i.e., when $\theta \in \Theta_a(r_\varepsilon)$ and $\xi \in \Xi$, where Ξ is defined in (1.2)); one would like to ensure that it is bounded by a prescribed level $\beta \in]0, 1[$.

We emphasize that in the classical minimax signal detection problem, the protection against all possible noise distributions (i.e., $\sup_{\xi \in \Xi}$) is not required, since the noise distribution is completely known. However, in the more general setting that we consider, in order to produce some kind of robustness, we have adapted the definitions of type I and type II error probabilities to accommodate the (possible) uncertainty on the noise.

Let $\alpha, \beta \in]0, 1[$ be given, and let Ψ_α be an α-level test, i.e., Ψ_α is such that $\sup_{\xi \in \Xi} \mathbb{P}_{0,\xi}(\Psi_\alpha = 1) \leq \alpha$.

Definition 2.1. The separation radius of the α-level test Ψ_α over the class $\mathcal{E}_a$ is defined as

$$r_\varepsilon(\mathcal{E}_a, \Psi_\alpha, \beta) := \inf \bigg\{ r_\varepsilon > 0 \colon \sup_{\substack{\theta \in \Theta_a(r_\varepsilon) \\ \xi \in \Xi}} \mathbb{P}_{\theta,\xi}(\Psi_\alpha = 0) \leq \beta \bigg\}.$$

In some sense, the separation radius $r_\varepsilon(\mathcal{E}_a, \Psi_\alpha, \beta)$ corresponds to the smallest possible value of the available signal $\|\theta\|$ for which H_0 and H_1 can be 'separated' by the α-level test Ψ_α with prescribed type I and type II error probabilities, α and β, respectively.

Definition 2.2. The minimax separation radius $\tilde{r}_\varepsilon := \tilde{r}_\varepsilon(\mathcal{E}_a, \alpha, \beta) > 0$ over the class $\mathcal{E}_a$ is defined as

$$\tilde{r}_\varepsilon := \inf_{\tilde{\Psi}_\alpha} r_\varepsilon(\mathcal{E}_a, \tilde{\Psi}_\alpha, \beta), \tag{2.5}$$

where the infimum is taken over all α-level tests $\tilde{\Psi}_\alpha$.

The minimax separation radius $\tilde{r}_\varepsilon$ corresponds to the smallest radius $r_\varepsilon > 0$ such that there exists some α-level test $\tilde{\Psi}_\alpha$ for which the type II error probability is not greater than β.

It is worth mentioning that Definitions 2.1 and 2.2 are valid for any *fixed* $\varepsilon > 0$ (i.e., it is not required that $\varepsilon \to 0$). The performance of any given test Ψ_α is easy to handle in the sense that the type I error probability is bounded by α (i.e., Ψ_α is an α-level test) and that the dependence of the minimax separation radius $\tilde{r}_\varepsilon$ with respect to given α and β can be precisely described.

3. CONTROL OF THE UPPER AND LOWER BOUNDS

3.1. The Spectral Cut-Off Test and Control of the Upper Bound

We define below a spectral cut-off test for the SM model (1.1) with $\xi \in \Xi$, where Ξ is defined in (1.2). First, we show that it is an α-level test and then we obtain an upper bound for its type II error probability.

Given a bandwidth $D \in \mathcal{N}$ and $\alpha \in]0, 1[$, we consider the following spectral cut-off test

$$\Psi_{\alpha,D} := \mathbf{1}_{\{T_D \geq t_{1-\alpha,D}\}}, \tag{3.1}$$

where

$$T_D = \sum_{k=1}^{D} b_k^{-2}(y_k^2 - \varepsilon^2)$$

and $t_{1-\alpha,D}$ denotes a threshold depending on α and D. It is easily seen that, for all $D \in \mathcal{N}$,

$$\mathbb{E}_{\theta,\xi}[T_D] = \sum_{k=1}^{D} \theta_k^2,$$

and

$$\mathrm{Var}_{0,\xi}(T_D) = \underbrace{\varepsilon^4 \sum_{k=1}^{D} b_k^{-4} \mathbb{E}[(\xi_k^2 - 1)^2]}_{:=R_0(D)} + \underbrace{\varepsilon^4 \sum_{\substack{k,l=1 \\ k \neq l}}^{D} b_k^{-2} b_l^{-2} \mathbb{E}[(\xi_k^2 - 1)(\xi_l^2 - 1)]}_{:=S_0(D)}, \tag{3.2}$$

where the assumption $\xi \in \Xi$ guarantees that the above variance is finite for every $D \in \mathcal{N}$.

Proposition 3.1. *Let $\alpha \in]0, 1[$ be given. Consider the spectral cut-off test $\Psi_{\alpha,D}$ defined in* (3.1). *Then, for all $\varepsilon > 0$,*

$$\sup_{\xi \in \Xi} \mathbb{P}_{0,\xi}(\Psi_{\alpha,D} = 1) \leq \alpha$$

as soon as

$$t_{1-\alpha,D} \geq \frac{1}{\sqrt{\alpha}} \sqrt{R_0(D) + S_0(D)}. \tag{3.3}$$

The proof of this proposition is postponed to Section 7.1.

Remarks:

- Using simple bounds, it is easily seen that

$$R_0(D) + S_0(D) \leq C_1\varepsilon^4 \sum_{1\leq k\leq D} b_k^{-4} + C_1\varepsilon^4 \Big(\sum_{1\leq k\leq D} b_k^{-2} \Big)^2 \leq 2C_1\varepsilon^4 \Big(\sum_{1\leq k\leq D} b_k^{-2} \Big)^2, \qquad (3.4)$$

  where

$$C_1 := \sup_{\xi\in\Xi} \sup_{k\in\mathcal{N}} \mathbb{E}[(\xi_k^2 - 1)^2] < +\infty, \qquad (3.5)$$

  since

$$\sum_{k=1}^{D} b_k^{-4} \leq (\max_{1\leq k\leq D} b_k^{-2}) \sum_{k=1}^{D} b_k^{-2} \leq \Big(\sum_{k=1}^{D} b_k^{-2} \Big)^2.$$

  Hence, the choice

$$t_{1-\alpha,D} = K_1\varepsilon^2 \sum_{k=1}^{D} b_k^{-2}, \qquad \text{where} \quad K_1 = \frac{\sqrt{2C_1}}{\sqrt{\alpha}}, \qquad (3.6)$$

  ensures that (3.3) is satisfied and that the spectral cut-off test $\Psi_{\alpha,D}$ defined in (3.1) is an α-level test.

- In the classical setting (i.e., independent Gaussian noise), the threshold $t_{1-\alpha,D}$ can be chosen as the $(1-\alpha)$-quantile of the variable T_D under H_0. This is no more the case here since only a uniform bound on the fourth moment of the sequence ξ_k, $k\in\mathcal{N}$, is available.

Proposition 3.2. *Let $\alpha,\beta \in\,]0,1[$ be given. Consider the spectral cut-off test $\Psi_{\alpha,D}$ defined in* (3.1). *Select the threshold $t_{1-\alpha,D}$ as in* (3.6). *Then, for all $\varepsilon > 0$,*

$$\sup_{\substack{\theta\in\Theta_a(r_{\varepsilon,D}) \\ \xi\in\Xi}} \mathbb{P}_{\theta,\xi}(\Psi_{\alpha,D} = 0) \leq \beta,$$

for any radius $r_{\varepsilon,D} > 0$ such that

$$r_{\varepsilon,D} \geq \mathcal{C}_\beta\varepsilon^2 \sum_{k=1}^{D} b_k^{-2} + a_D^{-2},$$

where $\mathcal{C}_\beta > 0$ is the solution of the equation (7.11).

The proof of Proposition is postponed to Section 7.2.

Remark. For practical purposes, the solution $\mathcal{C}_\beta > 0$ of equation (7.11) can be chosen as $\mathcal{C}_\beta = 8K_2/\beta$. In particular, there exists some $\beta_0 > 0$ such that $1 - K_1\mathcal{C}_\beta^{-1} \geq 1/2$ for all $\beta \leq \beta_0$, hence ensuring that (7.9) is satisfied for all β small enough.

3.2. Control of the Lower Bound

We propose below a lower bound on the minimax type II error probability for the SM (1.1) with $\xi \in \Xi$, where Ξ is defined in (1.2). In the sequel, the term $\inf_{\Psi_\alpha}$ corresponds to an infimum taken over all possible α-level tests.

Proposition 3.3. *Let $\alpha \in\,]0,1[$ and $\beta \in\,]0,1-\alpha[$ be fixed. Then, for all $\varepsilon > 0$*

$$\inf_{\Psi_\alpha} \sup_{\substack{\theta\in\Theta_a(r_\varepsilon) \\ \xi\in\Xi}} \mathbb{P}_{\theta,\xi}(\Psi_\alpha = 0) \geq \beta,$$

for all $D \in \mathcal{N}$ and $r_\varepsilon > 0$ such that

$$r_\varepsilon^2 \leq \left(\frac{1}{4}\log(\mathcal{C}_{\alpha,\beta})\right)\varepsilon^2 \sum_{k=1}^{D} b_k^{-2} \wedge a_D^{-2},$$

where $\mathcal{C}_{\alpha,\beta} = 1 + 4(1 - \alpha - \beta)^2$.

The proof of Proposition 3.3 is postponed to Section 7.3. The main difficulty is to construct an appropriate distribution for ξ that will allow one to obtain the largest possible lower bound.

4. MINIMAX SEPARATION RADIUS

The following theorem provides upper and lower bounds for the minimax separation radius $\tilde{r}_\varepsilon > 0$ in the SM (1.1) with $\xi \in \Xi$, where Ξ is defined in (1.2).

Theorem 4.1. *Let $\alpha, \beta \in \,]0, 1[$ be given. Then, for all $\varepsilon > 0$, the minimax separation radius $\tilde{r}_\varepsilon > 0$ satisfies*

$$\sup_{D \in \mathcal{N}}\left[\left(\frac{1}{4}\log(\mathcal{C}_{\alpha,\beta})\right)\varepsilon^2 \sum_{k=1}^{D} b_k^{-2} \wedge a_D^{-2}\right] \leq \tilde{r}_\varepsilon^2 \leq \inf_{D \in \mathcal{N}}\left[\mathcal{C}_\beta \varepsilon^2 \sum_{k=1}^{D} b_k^{-2} + a_D^{-2}\right], \qquad (4.1)$$

where $\mathcal{C}_\beta > 0$ is the solution of the equation (7.11) and $\mathcal{C}_{\alpha,\beta} = 1 + 4(1 - \alpha - \beta)^2$.

The proof of Theorem 4.1 is postponed to Section 7.4.

Remark. If both sequences $a = (a_k)_{k \in \mathcal{N}}$ and $b^{-1} = (b_k^{-1})_{k \in \mathcal{N}}$ are non-decreasing and satisfy

$$a_\star \leq \frac{a_{D-1}}{a_D} \leq a^\star \qquad \text{and} \qquad b_\star \leq \frac{b_{D-1}}{b_D} \leq b^\star \qquad \text{for all} \quad D \in \mathcal{N} \setminus \{1\}, \qquad (4.2)$$

for some constants $0 < a_\star \leq a^\star < \infty$ and $0 < b_\star \leq b^\star < \infty$, then it is easily seen that both upper and lower bounds on the minimax separation radius $\tilde{r}_\varepsilon > 0$ established in Theorem 4.1 are of the same order. This follows easily working along the same lines as in the proof of Proposition 4.1 in [15]. We note also that condition (4.2) is satisfied for various combinations of interest, among them: (i) mildly ill-posed inverse problems ($b_k \asymp k^{-t}$, $k \in \mathbb{N}$, for some $t > 0$) with ordinary smooth functions ($a_k \asymp k^s$, $k \in \mathbb{N}$, for some $s > 0$), (ii) severely ill-posed inverse problems ($b_k \asymp e^{-kt}$, $k \in \mathbb{N}$, for some $t > 0$) with ordinary smooth functions ($a_k \asymp k^s$, $k \in \mathbb{N}$, for some $s > 0$), and (iii) mildly ill-posed inverse problems ($b_k \asymp k^{-t}$, $k \in \mathbb{N}$, for some $t > 0$) with super-smooth functions ($a_k \asymp e^{ks}$, $k \in \mathbb{N}$, for some $s > 0$). Among the possible situations where condition (4.2) is not satisfied, one can mention, for instance, power-exponential behavior ($a_k \asymp e^{k^l s}$, $j \in \mathbb{N}$, for some $s > 0$ and $l > 1$, or $b_k \asymp e^{-k^r t}$, $k \in \mathbb{N}$, for some $t > 0$ and $r > 1$). See also Remark 4.3 in [15].

Remark. Note that the upper and lower bounds on the minimax separation radius $\tilde{r}_\varepsilon > 0$ established in Theorem 4.1 are quite different from the classical minimax separation radii available in the literature, obtained in the SM (1.1) with independent standard Gaussian noise (see, e.g., [15]). Although the bias terms a_D^{-2} coincide, the corresponding variance terms differ. In particular, in the SM (1.1) with $\xi \in \Xi$, where Ξ is defined in (1.2), the variance term is of order $\varepsilon^2 \sum_{k=1}^{D} b_k^{-2}$, while for the SM (1.1) with independent standard Gausiian noise, the variance term is of order $\varepsilon^2 \sqrt{\sum_{k=1}^{D} b_k^{-4}}$. We stress that the term $\varepsilon^2 \sum_{k=1}^{D} b_k^{-2}$ is not greater than the term $\varepsilon^2 \sqrt{\sum_{k=1}^{D} b_k^{-4}}$, which entails that the minimax separation rates are not faster compared to the ones obtained in the classical model. It is also worth mentioning that, surprisingly, the bias and variance terms in the SM (1.1) with $\xi \in \Xi$, where Ξ is defined in (1.2), are of the same order as the corresponding terms in the classical minimax estimation setting. In particular, the minimax separation rates in our general setting coincide with the minimax estimation rates obtained in the classical estimation setting. For illustrative purposes, Table 1 (see also Table 1 in [5]) provides these minimax separation rates for benchmark problems, i.e., well-posed, mildly ill-posed and severely ill-posed problems for ellipsoids with ordinary smooth and super-smooth sequences.

Table 1. Minimax separation rates for the SM (1.1) with $\xi \in \Xi$, where Ξ is defined in (1.2).

Minimax separation rate ($\tilde{r}_\varepsilon^2$)	ordinary-smooth $a_k \sim k^s$	super-smooth $a_k \sim \exp\{ks\}$
well-posed $b_k \sim 1$	$\varepsilon^{4s/(2s+1)}$	$\varepsilon^2(\log \varepsilon^{-1})$
mildly ill-posed $b_k \sim k^{-t}$	$\varepsilon^{4s/(2s+2t+1)}$	$\varepsilon^2(\log \varepsilon^{-1})^{2t+1}$
severely ill-posed $b_k \sim \exp\{-kt\}$	$(\log \varepsilon^{-1})^{-2s}$	$\varepsilon^{4s/(2s+2t)}$

Remark. If the supremum over all possible noise distributions $\xi \in \Xi$ is not considered in the definition of type I and type II error probabilities, then it is easily seen that the upper bound on the type II error probability obtained in Proposition 3.2 still holds true. However, the corresponding lower bound obtained in Proposition 3.3 is only true under Gaussianity. This implies that the minimax separation rates displayed in Table 1 are still valid in the SM (1.1) with non-independent standard Gaussian noise ξ.

5. AN ADDITIONAL CONDITION ON THE NOISE TO OBTAIN THE CLASSICAL MINIMAX SEPARATION RATES

In this section, it is demonstrated that, under an additional condition on the noise $\xi \in \Xi$ in the SM (1.1), one is able to retrieve the classical minimax separation rates in benchmark well-posed and ill-posed inverse problems.

Recall from equation (7.2), displayed in the proof of Proposition 3.2, that the variance of T_D can be written as

$$\mathrm{Var}_{\theta,\xi}(T_D) = R_\theta(D) + S_\theta(D),$$

where

$$R_\theta(D) := \sum_{k=1}^{D} b_k^{-4} \mathrm{Var}_{\theta,\xi}(y_k^2 - \varepsilon^2)$$

and

$$S_\theta(D) := \sum_{\substack{k,l=1 \\ k \neq l}}^{D} b_k^{-2} b_l^{-2} \mathrm{Cov}_{\theta,\xi}(y_k^2 - \varepsilon^2, y_l^2 - \varepsilon^2).$$

In the classical setting (i.e., independent standard Gaussian noise ξ), $S_\theta(D) = 0$ for all $\theta \in l^2(\mathcal{N})$. Hence in order to retrieve the classical minimax separation rates in the SM (1.1) with $\xi \in \Xi$, where Ξ is defined in (1.2), $S_\theta(D)$ needs to be of the order of $R_\theta(D)$. We achieve this separately under the null and the alternative hypotheses, for benchmark problems, such as well-posed, mildly ill-posed and severely ill-posed inverse problems.

We stress that in this section we will only deal with upper bounds. Indeed, the lower bounds established previously in the literature (see, e.g., [15], Theorem 4.1) for the independent standard Gaussian noise are still valid in our set-up.

5.1. Well-Posed and Mildly Ill-Posed Inverse Problems

We assume that

$$b_k \sim k^{-t} \quad \forall k \in \mathcal{N}$$

for some $t \geq 0$ ($t = 0$ refers to well-posed inverse problems while $t > 0$ refers to mildly ill-posed inverse problems). We start our discussion under the null hypothesis. Recall from (3.2) that

$$\mathrm{Var}_{0,\xi}(T_D) = \varepsilon^4 \sum_{k=1}^{D} b_k^{-4} \mathbb{E}[(\xi_k^2 - 1)^2] + \varepsilon^4 \sum_{\substack{k,l=1 \\ k \neq l}}^{D} b_k^{-2} b_l^{-2} \mathbb{E}[(\xi_k^2 - 1)(\xi_l^2 - 1)]$$

$$= \varepsilon^4 \sum_{k=1}^{D} b_k^{-4} \mathrm{Var}(\xi_k^2) + \varepsilon^4 \sum_{\substack{k,l=1 \\ k \neq l}}^{D} b_k^{-2} b_l^{-2} \mathrm{Cov}(\xi_k^2, \xi_l^2)$$

$$:= R_0(D) + S_0(D).$$

Using simple calculations, we can see that

$$R_0(D) \sim \varepsilon^4 \sum_{k=1}^{D} k^{4t} \sim \varepsilon^4 D^{4t+1}.$$

Our aim is to exhibit a condition for which $S_0(D)$ is (at least) of the same order as $R_0(D)$.

Assumption $\mathcal{H}_{\mathcal{D}}$: *Let $\xi \in \Xi$, where Ξ is defined in (1.2), and, for all $k, l \in \mathcal{N}$, let $(\xi_k, \xi_l)'$ be a bivariate Gaussian random vector. Moreover, there exists $s > 0$ such that*

$$\rho_{kl} := |\mathrm{Cov}(\xi_k, \xi_l)| \lesssim \frac{1}{|k - l|^s} \quad \forall k, l \in \mathcal{N}, \ k \neq l.$$

Due to the Isserlis Theorem (see, e.g., [9]), it can be seen that, thanks to Assumption $\mathcal{H}_{\mathcal{D}}$, for all $k, l \in \mathcal{N}$ with $k \neq l$,

$$\mathrm{Cov}(\xi_k^2, \xi_l^2) = 2\mathrm{Cov}^2(\xi_k, \xi_l) \lesssim \frac{1}{|k - l|^{2s}} \tag{5.1}$$

and

$$\mathbb{E}[(\xi_k^2 - 1)\xi_l] = \mathbb{E}[(\xi_l^2 - 1)\xi_k] = 0. \tag{5.2}$$

These results allow us to propose a sharp control of the variance of T_D under the null hypothesis.

Proposition 5.1. *Assume that Assumption $\mathcal{H}_{\mathcal{D}}$ holds with $s > 1/2$. Then*

$$S_0(D) = o(R_0(D)) \quad as \quad D \to +\infty.$$

The proof of Proposition 5.1 is postponed to Section 7.5.

Now, we propose a similar analysis under the alternative hypothesis.

Proposition 5.2. *Assume that Assumption $\mathcal{H}_{\mathcal{D}}$ holds with $s > 1/2$. Then, for all $\gamma \in]0, 1[$,*

$$\mathrm{Var}_{\theta,\xi}(T_D) \lesssim (1 + \gamma^{-1})\varepsilon^4 \sum_{k=1}^{D} b_k^{-4} + \gamma \Big(\sum_{k=1}^{D} \theta_k^2 \Big)^2.$$

The proof of Proposition 5.2 is postponed to Section 7.6.

Starting from (7.1) and using Propositions 5.1 and 5.2, we get

$$\mathbb{P}_{\theta,\xi}(\Psi_{\alpha,D} = 0) \leq \frac{\mathrm{Var}_{\theta,\xi}(T_D)}{\left(\sum_{k=1}^{D} \theta_k^2 - t_{1-\alpha,D} \right)^2}$$

$$\lesssim \frac{(1 + \gamma^{-1})\varepsilon^4 \sum_{k=1}^{D} b_k^{-4} + \gamma \left(\sum_{k=1}^{D} \theta_k^2 \right)^2}{\left(\sum_{k=1}^{D} \theta_k^2 - \varepsilon^4 \sum_{k=1}^{D} b_k^{-4} \right)^2} \leq \beta$$

provided

$$\sum_{k=1}^{D} \theta_k^2 \gtrsim \varepsilon^2 \sqrt{\sum_{k=1}^{D} b_k^{-4}},$$

which holds as soon as

$$\|\theta\|^2 \gtrsim a_D^{-2} + \varepsilon^2 \sqrt{\sum_{k=1}^{D} b_k^{-4}}.$$

The last inequality provides a classical condition that has been already discussed in, e.g., [1], [7] and [12], or in Theorem 4.1 of [15], in the specific case where the noise ξ in the SM (1.1) is assumed to be independent standard Gaussian. This entails that the Assumption $\mathcal{H}_{\mathcal{D}}$ suffices to retrieve the classical minimax separation rates for mildly ill-posed inverse problems.

5.2. Severely Ill-Posed Inverse Problems

We assume in this section that

$$b_k \sim e^{-kt} \quad \forall k \in \mathcal{N}$$

for some $t > 0$. Since minimax estimation and minimax separation rates in the classical setting are of the same order (see, e.g., Tables 2 and 3 in [7]), we stress that non-independence does not deteriorate the classical minimax separation rates. In other words, the independent standard Gaussian assumption on noise ξ is not needed to get the classical minimax separation rates for severely ill-posed inverse problems.

6. CONCLUDING REMARKS

We have established minimax separation rates in a general Gaussian sequence model, i.e., the noise need neither to be independent nor standard Gaussian. These rates are not faster than the ones obtained in the classical setting (i.e., independent standard Gaussian noise) but, surprisingly, are of the same order as the minimax estimation rates in the classical setting. The involved spectral cut-off test depends on the unknown smoothness parameter of the signal under the alternative hypothesis. It is therefore of paramount importance in practical applications to provide minimax testing procedures that do not explicitly depend on the associated smoothness parameter. This is, usually, referred to as the 'adaptation' problem. However, such an investigation needs careful attention that is beyond the scope of the present work. In particular, the dependence of the involved constant with respect to the level α has a more intricate form than the one involved in the classical setting.

7. APPENDIX

7.1. Proof of Proposition 3.1

Let $\xi \in \Xi$ be fixed. Using the Markov inequality, we get

$$\mathbb{P}_{0,\xi}(\Psi_{\alpha,D} = 1) = \mathbb{P}_{0,\xi}(T_D \geq t_{1-\alpha,D}) \leq \frac{\mathrm{Var}_{0,\xi}(T_D)}{t_{1-\alpha,D}^2} \leq \frac{R_0(D) + S_0(D)}{t_{1-\alpha,D}^2} \leq \alpha$$

provided

$$t_{1-\alpha,D} \geq \frac{1}{\sqrt{\alpha}} \sqrt{R_0(D) + S_0(D)}.$$

$\square$

7.2. Proof of Proposition 3.2

Let $\xi \in \Xi$ be fixed. Using the Markov inequality, we obtain

$$\mathbb{P}_{\theta,\xi}(\Psi_{\alpha,D} = 0) = \mathbb{P}_{\theta,\xi}(T_D < t_{1-\alpha,D}) = \mathbb{P}_{\theta,\xi}\Big(T_D - \mathbb{E}_\theta[T_D] < t_{1-\alpha,D} - \sum_{k=1}^{D}\theta_k^2\Big)$$

$$\leq \mathbb{P}_{\theta,\xi}\Big(\big|T_D - \mathbb{E}_\theta[T_D]\big| \geq \sum_{k=1}^{D}\theta_k^2 - t_{1-\alpha,D}\Big)$$

$$\leq \frac{\mathrm{Var}_{\theta,\xi}(T_D)}{\big(\sum_{k=1}^{D}\theta_k^2 - t_{1-\alpha,D}\big)^2}, \tag{7.1}$$

where we have implicitly assumed that

$$\sum_{k=1}^{D}\theta_k^2 > t_{1-\alpha,D}.$$

Now, we need an upper bound for the variance term. First remark that

$$\mathrm{Var}_\theta(T_D) = \mathrm{Var}_{\theta,\xi}\Big(\sum_{k=1}^{D}b_k^{-2}(y_k^2 - \varepsilon^2)\Big)$$

$$= \underbrace{\sum_{k=1}^{D}b_k^{-4}\mathrm{Var}_{\theta,\xi}(y_k^2 - \varepsilon^2)}_{:=R_\theta(D)} + \underbrace{\sum_{\substack{k,l=1 \\ k \neq l}}^{D}b_k^{-2}b_l^{-2}\mathrm{Cov}_{\theta,\xi}(y_k^2 - \varepsilon^2, y_l^2 - \varepsilon^2)}_{:=S_\theta(D)}. \tag{7.2}$$

Calculation of $R_\theta(D)$: Using simple algebra, we get, for all $k \in \mathcal{N}$,

$$\mathrm{Var}_\theta(y_k^2 - \varepsilon^2) = \mathrm{Var}_{\theta,\xi}\big[(b_k\theta_k + \varepsilon)^2 - \varepsilon^2\big]$$

$$= \mathrm{Var}_{\theta,\xi}\big[b_k^2\theta_k^2 + \varepsilon^2\xi_k^2 + 2b_k\theta_k\varepsilon\xi_k - \varepsilon^2\big]$$

$$= \mathrm{Var}_{\theta,\xi}\big[\varepsilon^2(\xi_k^2 - 1) + 2b_k\theta_k\varepsilon\xi_k\big]$$

$$= \varepsilon^4\mathbb{E}[(\xi_k^2 - 1)^2] + 4\varepsilon^2 b_k^2\theta_k^2 + 4\varepsilon^3 b_k\theta_k\mathbb{E}[\xi_k^3].$$

Hence, using the last equality, we obtain

$$R_\theta(D) \leq C_1\varepsilon^4\sum_{k=1}^{D}b_k^{-4} + 4\varepsilon^2\sum_{k=1}^{D}b_k^{-2}\theta_k^2 + 4C_2\varepsilon^3\sum_{k=1}^{D}b_k^{-3}|\theta_k|$$

$$\leq C_1\varepsilon^4\sum_{k=1}^{D}b_k^{-4} + 4\varepsilon^2\big(\max_{1 \leq k \leq D}b_k^{-2}\big)\sum_{k=1}^{D}\theta_k^2 + 4C_2\varepsilon^3\sum_{k=1}^{D}b_k^{-3}|\theta_k|, \tag{7.3}$$

where the constant C_1 has been introduced in (3.5) and

$$C_2 := \sup_{\xi \in \Xi}\sup_{k \in \mathcal{N}}|\mathbb{E}[\xi_k^3]| < +\infty.$$

Note that, using first the Cauchy–Schwarz inequality and then the Peter-Paul inequality (see, e.g., [3], p. 18), we get

$$\varepsilon^3\sum_{k=1}^{D}b_k^{-3}|\theta_k| = \sum_{k=1}^{D}\varepsilon^2 b_k^{-2}\varepsilon b_k^{-1}|\theta_k| \leq \sqrt{\varepsilon^4\sum_{k=1}^{D}b_k^{-4}}\sqrt{\varepsilon^2\sum_{k=1}^{D}b_k^{-2}\theta_k^2}$$

$$\leq \frac{1}{2}\varepsilon^4\sum_{k=1}^{D}b_k^{-4} + \frac{1}{2}\varepsilon^2\sum_{k=1}^{D}b_k^{-2}\theta_k^2 \leq \frac{1}{2}\varepsilon^4\sum_{k=1}^{D}b_k^{-4} + \frac{1}{2}\varepsilon^2\big(\max_{1 \leq k \leq D}b_k^{-2}\big)\sum_{k=1}^{D}\theta_k^2. \tag{7.4}$$

Combining inequalities (7.3) and (7.4), we obtain

$$R_\theta(D) \leq (C_1 + 2C_2)\varepsilon^4 \sum_{k=1}^{D} b_k^{-4} + (4 + 2C_2)\varepsilon^2 (\max_{1 \leq k \leq D} b_k^{-2}) \sum_{k=1}^{D} \theta_k^2. \tag{7.5}$$

Calculation of $S_\theta(D)$: First, remark that for all $k \in \mathcal{N}$, on noting that $\mathbb{E}_{\theta,\xi}[(y_k^2 - \varepsilon^2)] = b_k^2\theta_k^2$ and $\mathbb{E}[\xi_k^2] = 1$, we get

$$\begin{aligned}
&\mathrm{Cov}_{\theta,\xi}(y_k^2 - \varepsilon^2, y_l^2 - \varepsilon^2)\\
&= \mathbb{E}_{\theta,\xi}\big[(y_k^2 - \varepsilon^2 - b_k^2\theta_k^2)(y_l^2 - \varepsilon^2 - b_l^2\theta_l^2)\big]\\
&= \mathbb{E}_{\theta,\xi}\big[((b_k\theta_k + \varepsilon\xi_k)^2 - \varepsilon^2 - b_k^2\theta_k^2)((b_l\theta_l + \varepsilon\xi_l)^2 - \varepsilon^2 - b_l^2\theta_l^2)\big]\\
&= \varepsilon^4 \mathbb{E}[(\xi_k^2 - 1)(\xi_l^2 - 1)] + 4\varepsilon^2 b_k b_l \theta_k \theta_l \mathbb{E}[\xi_k\xi_l] + 2\varepsilon^3 b_l \theta_l \mathbb{E}[(\xi_k^2 - 1)\xi_l] + 2\varepsilon^3 b_k \theta_k \mathbb{E}[(\xi_l^2 - 1)\xi_k].
\end{aligned}$$

Hence

$$\begin{aligned}
S_\theta(D) = {}&2\varepsilon^4 \sum_{\substack{k,l=1\\k\neq l}}^{D} b_k^{-2}b_l^{-2}\mathbb{E}[(\xi_k^2 - 1)(\xi_l^2 - 1)] + 8\varepsilon^2 \sum_{\substack{k,l=1\\k\neq l}}^{D} b_k^{-1}b_l^{-1}\theta_k\theta_l\mathbb{E}[\xi_k\xi_l]\\
&+ 4\varepsilon^3 \sum_{\substack{k,l=1\\k\neq l}}^{D} b_k^{-2}b_l^{-1}\theta_l\mathbb{E}[(\xi_k^2 - 1)\xi_l] + 4\varepsilon^3 \sum_{\substack{k,l=1\\k\neq l}}^{D} b_l^{-2}b_k^{-1}\theta_k\mathbb{E}[(\xi_l^2 - 1)\xi_k].
\end{aligned} \tag{7.6}$$

Using the Cauchy–Schwarz inequality in each expectation of the above expression we obtain

$$S_\theta(D) \leq 2C_1\varepsilon^4 \Big(\sum_{k=1}^{D} b_k^{-2}\Big)^2 + 8\varepsilon^2 \Big(\sum_{k=1}^{D} b_k^{-1}|\theta_k|\Big)^2 + 8\varepsilon^3 C_1^{1/2}\Big(\sum_{k=1}^{D} b_k^{-1}|\theta_k|\Big)\Big(\sum_{k=1}^{D} b_k^{-2}\Big).$$

Then, using first the Peter-Paul inequality and then the Cauchy–Schwarz inequality, we get

$$\begin{aligned}
S_\theta(D) &\leq 2(C_1 + 2C_1^{1/2})\varepsilon^4 \Big(\sum_{k=1}^{D} b_k^{-2}\Big)^2 + 4(2 + C_1^{1/2})\varepsilon^2 \Big(\sum_{k=1}^{D} b_k^{-1}|\theta_k|\Big)^2\\
&\leq 2(C_1 + 2C_1^{1/2})\varepsilon^4 \Big(\sum_{k=1}^{D} b_k^{-2}\Big)^2 + 4(2 + C_1^{1/2})\varepsilon^2 \sum_{k=1}^{D} b_k^{-2} \sum_{k=1}^{D} \theta_k^2.
\end{aligned} \tag{7.7}$$

Hence combining (7.2), (7.5) and (7.7) we obtain

$$\begin{aligned}
\mathrm{Var}_{\theta,\xi}(T_D) \leq {}&(C_1 + 2C_2)\varepsilon^4 \sum_{k=1}^{D} b_k^{-4} + 4(C_1 + 2C_1^{1/2})\varepsilon^4 \Big(\sum_{k=1}^{D} b_k^{-2}\Big)^2\\
&+ 2(10 + C_2 + 4C_1^{1/2})\varepsilon^2 \sum_{k=1}^{D} b_k^{-2} \sum_{k=1}^{D} \theta_k^2.
\end{aligned}$$

For all $\gamma \in]0, 1[$, using again the Peter-Paul inequality, we get

$$\begin{aligned}
\mathrm{Var}_{\theta,\xi}(T_D) \leq {}&(C_1 + 2C_2)\varepsilon^4 \sum_{k=1}^{D} b_k^{-4} + 4(C_1 + 2C_1^{1/2})\varepsilon^4 \Big(\sum_{k=1}^{D} b_k^{-2}\Big)^2\\
&+ \gamma^{-1}(10 + C_2 + 4C_1^{1/2})\varepsilon^4 \Big(\sum_{k=1}^{D} b_k^{-2}\Big)^2 + \gamma(10 + C_2 + 4C_1^{1/2})\Big(\sum_{k=1}^{D} \theta_k^2\Big)^2.
\end{aligned}$$

Hence, since $\gamma \in]0, 1[$, it is easily seen that

$$
\mathrm{Var}_{\theta,\xi}(T_D) \le \gamma(10 + C_2 + 4C_1^{1/2})\Big(\sum_{k=1}^{D} \theta_k^2\Big)^2
$$

$$
+ \Big\{(C_1 + 2C_2) + 4(C_1 + 2C_1^{1/2}) + \gamma^{-1}(10 + C_2 + 4C_1^{1/2})\Big\}\varepsilon^4\Big(\sum_{k=1}^{D} b_k^{-2}\Big)^2
$$

$$
\le \gamma(10 + C_2 + 4C_1^{1/2})\Big(\sum_{k=1}^{D} \theta_k^2\Big)^2 + \gamma^{-1}\big\{10 + 5C_1 + 2C_2 + 12C_1^{1/2}\big\}\varepsilon^4\Big(\sum_{k=1}^{D} b_k^{-2}\Big)^2
$$

$$
\le (10 + 5C_1 + 2C_2 + 12C_1^{1/2})\Big\{\gamma\Big(\sum_{k=1}^{D} \theta_k^2\Big)^2 + \gamma^{-1}\varepsilon^4\Big(\sum_{k=1}^{D} b_k^{-2}\Big)^2\Big\}
$$

$$
:= K_2\Big\{\gamma\Big(\sum_{k=1}^{D} \theta_k^2\Big)^2 + \gamma^{-1}\varepsilon^4\Big(\sum_{k=1}^{D} b_k^{-2}\Big)^2\Big\}. \tag{7.8}
$$

Now, using (7.1) and (7.8) and choosing $\gamma = \mathcal{C}_\beta^{-1}$, we get

$$
\mathbb{P}_{\theta,\xi}(\Psi_{\alpha,D} = 0) \le \frac{K_2\big\{\gamma(\sum_{k=1}^{D} \theta_k^2)^2 + \gamma^{-1}\varepsilon^4(\sum_{k=1}^{D} b_k^{-2})^2\big\}}{\big(\sum_{k=1}^{D} \theta_k^2 - K_1\varepsilon^2\sum_{k=1}^{D} b_k^{-2}\big)^2}
$$

$$
\le \frac{K_2\big\{\gamma + \gamma^{-1}\mathcal{C}_\beta^{-2}\big\}}{\big(1 - K_1\mathcal{C}_\beta^{-1}\big)^2} \le \frac{2K_2\mathcal{C}_\beta^{-1}}{\big(1 - K_1\mathcal{C}_\beta^{-1}\big)^2} \le \beta, \tag{7.9}
$$

provided that

$$
\sum_{k=1}^{D} \theta_k^2 \ge \mathcal{C}_\beta\varepsilon^2\sum_{k=1}^{D} b_k^{-2} \tag{7.10}
$$

and $\mathcal{C}_\beta$ is the solution of the equation

$$
\frac{2K_2\mathcal{C}_\beta^{-1}}{\big(1 - K_1\mathcal{C}_\beta^{-1}\big)^2} = \beta. \tag{7.11}
$$

To conclude the proof, since $\sum_{k>D} \theta_k^2 \le a_D^{-2}$, remark that inequality (7.10) is satisfied provided that

$$
\|\theta\|^2 \ge \mathcal{C}_\beta\varepsilon^2\sum_{k=1}^{D} b_k^{-2} + a_D^{-2}.
$$

$\square$

7.3. Proof of Proposition 3.3

When ξ is Gaussian, we will write Ξ_G instead of Ξ and $\mathbb{P}_{\theta,\Sigma}$ instead of $\mathbb{P}_{\theta,\xi}$, where $\Sigma = (\Sigma_{kl})_{k,l \in \mathcal{N}}$ denotes the associated covariance matrix. We also define $S = \{\Sigma \colon \Sigma_{kk} = 1\}$. Below, Ψ_α refers to an α-level test.

Let $\theta^\star \in \Theta_a(r_\epsilon)$, $\xi^\star \in \Xi_G$ and $\Sigma^\star \in S$ be fixed. Their values will be made precise later on. Then

$$
\inf_{\Psi_\alpha} \sup_{\substack{\theta \in \Theta_a(r_\epsilon) \\ \xi \in \Xi}} \mathbb{P}_{\theta,\xi}(\Psi_\alpha = 0) \ge \inf_{\Psi_\alpha} \sup_{\substack{\theta \in \Theta_a(r_\epsilon) \\ \xi \in \Xi_G}} \mathbb{P}_{\theta,\xi}(\Psi_\alpha = 0)
$$

$$
= \inf_{\Psi_\alpha} \sup_{\substack{\theta \in \Theta_a(r_\epsilon) \\ \tilde{\Sigma} \in S}} \mathbb{P}_{\theta,\tilde{\Sigma}}(\Psi_\alpha = 0) \ge \inf_{\Psi_\alpha} \sup_{\theta \in \Theta_a(r_\epsilon)} \mathbb{P}_{\theta,\Sigma^\star}(\Psi_\alpha = 0)
$$

$$\geq \inf_{\Psi_\alpha} \mathbb{P}_{\theta^\star,\Sigma^\star}(\Psi_\alpha = 0) \geq 1 - \alpha - \frac{1}{2}\big(\mathbb{E}_0[L^2_{\theta^\star,\Sigma^\star}(Y)] - 1\big)^{1/2},$$

where $L_{\theta^\star,\Sigma^\star}(Y) = d\mathbb{P}_{\theta^\star,\Sigma^\star}(Y)/d\mathbb{P}_{0,\Sigma^\star}(Y)$ is the likelihood ratio between the probability measures $\mathbb{P}_{\theta^\star,\Sigma^\star}$ and $\mathbb{P}_{0,\Sigma^\star}$ (for the last inequality, we refer to, e.g., (3.1) in [15]). In particular, if we can find $\theta^\star$ and $\Sigma^\star$ such that

$$\mathbb{E}_0[L^2_{\theta^\star,\Sigma^\star}(Y)] \leq \mathcal{C}_{\alpha,\beta}$$

for some $\beta \in]0, 1 - \alpha[$, then

$$\inf_{\Psi_\alpha} \sup_{\substack{\theta \in \Theta_a(r_\epsilon) \\ \xi \in \Xi}} \mathbb{P}_{\theta,\xi}(\Psi_\alpha = 0) \geq \beta.$$

Let $D \in \mathcal{N}$ be fixed. Now, we impose the following conditions on $\theta^\star$ and $\Sigma^\star$:

$$\theta^\star_k = 0 \quad \forall k > D \qquad \text{and} \qquad \Sigma^\star_{kl} = 0 \quad \forall k > D, \, l > D, \, k \neq l.$$

Let $\Sigma^\star_D = (\Sigma^\star_{kl})_{1 \leq k,l \leq D}$ be the remaining submatrix of $\Sigma^\star$. With a slight abuse of notation, we denote below $Y = (Y_1, \ldots, Y_D)'$ and $b\theta^\star = (b_1\theta^\star_1, \ldots, b_D\theta^\star_D)'$. Then, by simple algebra,

$$L_{\theta^\star,\Sigma^\star}(Y) = \frac{\exp\big(-\frac{1}{2\varepsilon^2}(Y - b\theta^\star)'(\Sigma^\star_D)^{-1}(Y - b\theta^\star)\big)}{\exp\big(-\frac{1}{2\varepsilon^2}Y'(\Sigma^\star_D)^{-1}Y\big)}$$

$$= \exp\Big[\frac{1}{2\varepsilon^2}\big\{2(b\theta^\star)'(\Sigma^\star_D)^{-1}Y - (b\theta^\star)'(\Sigma^\star_D)^{-1}b\theta^\star\big\}\Big].$$

Hence

$$\mathbb{E}_0[L^2_{\theta^\star,\Sigma^\star}(Y)] = \exp\Big(-\frac{1}{\varepsilon^2}(b\theta^\star)'(\Sigma^\star_D)^{-1}b\theta^\star\Big)\mathbb{E}_0\Big[\exp\Big(\frac{2}{\varepsilon^2}(b\theta^\star)'(\Sigma^\star_D)^{-1}Y\Big)\Big].$$

It is easily seen that

$$\mathbb{E}_0\Big[\exp\Big(\frac{2}{\varepsilon^2}(b\theta^\star)'(\Sigma^\star_D)^{-1}Y\Big)\Big]$$

$$= \frac{1}{(2\pi\varepsilon^2)^{D/2}|\Sigma^\star_D|^{1/2}} \int_{\mathbb{R}^D} \exp\Big(\frac{2}{\varepsilon^2}(b\theta^\star)'(\Sigma^\star_D)^{-1}y\Big) \exp\Big(-\frac{1}{2\varepsilon^2}y'(\Sigma^\star_D)^{-1}y\Big)\, dy$$

$$= \frac{1}{(2\pi\varepsilon^2)^{D/2}|\Sigma^\star_D|^{1/2}} \int_{\mathbb{R}^D} \exp\Big(-\frac{1}{2\epsilon^2}\big\{y'(\Sigma^\star_D)^{-1}y - 4(b\theta^\star)'(\Sigma^\star_D)^{-1}y\big\}\Big)\, dy$$

$$= \exp\Big(\frac{2}{\varepsilon^2}(b\theta^\star)'(\Sigma^\star_D)^{-1}b\theta^\star\Big)\frac{1}{(2\pi\varepsilon^2)^{D/2}|\Sigma^\star_D|^{1/2}} \int_{\mathbb{R}^D} \exp\Big(-\frac{1}{2\epsilon^2}(y - 2b\theta^\star)'(\Sigma^\star_D)^{-1}(y - 2b\theta^\star)\Big)\, dy$$

$$= \exp\Big(\frac{2}{\varepsilon^2}(b\theta^\star)'(\Sigma^\star_D)^{-1}b\theta^\star\Big).$$

Hence

$$\mathbb{E}_0[L^2_{\theta^\star,\Sigma^\star}(Y)] = \exp\Big(\frac{1}{\varepsilon^2}(b\theta^\star)'(\Sigma^\star_D)^{-1}b\theta^\star\Big).$$

Now, we select $\theta^\star$ as follows

$$\theta^\star_k = \frac{r_\varepsilon b^{-1}_k(\Sigma^\star_D v)_k}{\rho} \quad \forall k \in \{1, \ldots, D\} \qquad \text{and} \qquad \theta^\star_k = 0 \;\; \forall k > D,$$

where

$$\rho^2 = \sum_{k=1}^D b^{-2}_k(\Sigma^\star_D v)^2_k \qquad \text{and} \qquad v_k = \frac{1}{\sqrt{D}} \;\; \forall k \in \{1, \ldots, D\}. \tag{7.12}$$

Now, define $v = (v_1, \ldots, v_D)'$ and note that $\|v\| = 1$. Then it is easily seen that

$$\mathbb{E}_0[L^2_{\theta^\star,\Sigma^\star}(Y)] = \exp\Big(\frac{r^2_\varepsilon}{\varepsilon^2\rho^2}v'\Sigma^\star_D v\Big). \tag{7.13}$$

We first construct a specific $\xi^\star \in \Xi_G$. Let

$$\xi_k^\star = d_k \eta_0 + \sqrt{1 - d_k^2}\, \eta_k \quad \forall k \in \{1, \ldots, D\} \qquad \text{and} \qquad \xi_k^\star = \eta_k \quad \forall k > D, \qquad (7.14)$$

where $(\eta_k)_{k \in \mathcal{N}_0}$ denotes a sequence of independent standard Gaussian random variables and $d = (d_k)_{1 \leq k \leq D}$ is a real sequence such that $1/\sqrt{2} \leq d_k < 1$ for all $1 \leq k \leq D$. Obviously, $\xi^\star \in \Xi_G$, since $\xi^\star$ is Gaussian,

$$\mathbb{E}[\xi_k^\star] = 0, \qquad \mathbb{E}[(\xi_k^\star)^2] = 1 \quad \forall k \in \mathcal{N}$$

and

$$\max_{1 \leq k \leq D} \mathbb{E}[(\xi_k^\star)^4] \leq \max_{1 \leq k \leq D} \left[8d_k^4 \mathbb{E}[\xi_0^4] + 8(1 - d_k)^2 \mathbb{E}[\eta_k^4] \right] \leq 16\mathbb{E}[\xi_0^4] = C < +\infty,$$

$$\sup_{k > D} \mathbb{E}[(\xi_k^\star)^4] = 3.$$

Now, we need to bound the expression in (7.13). Using (7.14), we get

$$\Sigma_{kl}^\star = \mathbb{E}[\xi_k^\star \xi_l^\star] = d_k d_l \geq 1/2 \quad \forall k, l \in \{1, \ldots, D\} \text{ with } k \neq l \quad \text{and} \quad \Sigma_{kl} = 0 \quad \forall k, l > D, \ k \neq l.$$

Note also that, since v is a unit vector,

$$v' \Sigma_D^\star v \leq \max_{\|a\|=1} a' \Sigma_D^\star a \leq D$$

since the largest eigenvalue of $\Sigma_D^\star$ is smaller than D. Now, using (7.12), we get

$$\rho^2 = \sum_{k=1}^{D} b_k^{-2} (\Sigma_D^\star v)_k^2 = \sum_{k=1}^{D} \left(\sum_{l=1}^{D} \Sigma_{kl}^\star v_l \right)^2 \geq \frac{1}{4} D \sum_{k=1}^{D} b_k^{-2}$$

since $\Sigma_{kl}^\star \geq 1/2$ for all $k, l \in \{1, \ldots, D\}$. Hence

$$\mathbb{E}_0[L_{\theta^\star, \Sigma^\star}^2(Y)] \leq \exp\left(\frac{4r_\varepsilon^2}{\varepsilon^2} \frac{1}{\sum_{k=1}^{D} b_k^{-2}} \right) \leq \mathcal{C}_{\alpha, \beta}$$

provided

$$\|\theta\|^2 = r_\varepsilon^2 \leq \left(\frac{1}{4} \log(\mathcal{C}_{\alpha, \beta}) \right) \varepsilon^2 \sum_{k=1}^{D} b_k^{-2}.$$

To conclude the proof, we need to ensure that the constructed $\theta^\star$ belongs to $\mathcal{E}_a$. Remark that, since a is an increasing sequence,

$$\sum_{k \in \mathcal{N}} a_k^2 (\theta_k^\star)^2 = \sum_{k=1}^{D} a_k^2 (\theta_k^\star)^2 \leq a_D^2 \sum_{k=1}^{D} a_k^2 (\theta_k^\star)^2 \leq a_D^2 r_\varepsilon^2 \leq 1$$

provided $r_\varepsilon^2 \leq a_D^{-2}$. Hence

$$\inf_{\Psi_\alpha} \sup_{\substack{\theta \in \Theta_a(r_\varepsilon) \\ \xi \in \Xi}} \mathbb{P}_{\theta, \xi}(\Psi_\alpha = 0) \geq \beta$$

as soon as

$$r_\varepsilon^2 \leq \left(\frac{1}{4} \log(\mathcal{C}_{\alpha, \beta}) \right) \varepsilon^2 \sum_{k=1}^{D} b_k^{-2} \wedge a_D^{-2}.$$

$\square$

7.4. Proof of Theorem 4.1

In Proposition 3.2, we have proved that for all $D \in \mathcal{N}$ there exists an α-level test $\Psi_{\alpha,D}$ such that

$$\sup_{\substack{\theta \in \Theta_a(r_{\varepsilon,D}) \\ \xi \in \Xi}} \mathbb{P}_{\theta,\xi}(\Psi_{\alpha,D} = 0) \leq \beta,$$

for any radius $r_{\varepsilon,D} > 0$ satisfying

$$r_{\varepsilon,D}^2 \geq \mathcal{C}_\beta \varepsilon^2 \sum_{k=1}^{D} b_k^{-2} + a_D^{-2}.$$

Now, setting

$$D^\dagger = \arg \inf_{D \in \mathcal{N}} \left[\mathcal{C}_\beta \varepsilon^2 \sum_{k=1}^{D} b_k^{-2} + a_D^{-2} \right]$$

and denoting by $\Psi_{\alpha,D^\dagger}$ the associated α-level test, we get

$$\sup_{\substack{\theta \in \Theta_a(r_\varepsilon^\star) \\ \xi \in \Xi}} \mathbb{P}_{\theta,\xi}(\Psi_{\alpha,D^\dagger} = 0) \leq \beta$$

for any radius $r_\varepsilon^\star > 0$ satisfying

$$(r_\varepsilon^\star)^2 \geq \inf_{D \in \mathcal{N}} \left[\mathcal{C}_\beta \varepsilon^2 \sum_{k=1}^{D} b_k^{-2} + a_D^{-2} \right].$$

Hence

$$\tilde{r}_\varepsilon^2 \leq \inf_{D \in \mathcal{N}} \left[\mathcal{C}_\beta \varepsilon^2 \sum_{k=1}^{D} b_k^{-2} + a_D^{-2} \right].$$

Similarly, using Proposition 3.3,

$$\inf_{\Psi_\alpha} \sup_{\substack{\theta \in \Theta_a(r_\varepsilon) \\ \xi \in \Xi}} \mathbb{P}_{\theta,\xi}(\Psi_\alpha = 0) \geq \beta$$

for any radius $r_\varepsilon > 0$ such that

$$r_\varepsilon^2 \leq \left(\frac{1}{4} \log(\mathcal{C}_{\alpha,\beta}) \right) \varepsilon^2 \sum_{k=1}^{D} b_k^{-2} \wedge a_D^{-2}.$$

This result occurs for all $D \in \mathcal{N}$. Hence

$$\inf_{\Psi_\alpha} \sup_{\substack{\theta \in \Theta_a(r_\varepsilon) \\ \xi \in \Xi}} \mathbb{P}_{\theta,\xi}(\Psi_\alpha = 0) \geq \beta$$

for any radius $r_\varepsilon > 0$ such that

$$r_\varepsilon^2 \leq \sup_{D \in \mathcal{N}} \left[\left(\frac{1}{4} \log(\mathcal{C}_{\alpha,\beta}) \right) \varepsilon^2 \sum_{k=1}^{D} b_k^{-2} \wedge a_D^{-2} \right].$$

This entails that

$$\tilde{r}_\varepsilon^2 \geq \sup_{D \in \mathcal{N}} \left[\left(\frac{1}{4} \log(\mathcal{C}_{\alpha,\beta}) \right) \varepsilon^2 \sum_{k=1}^{D} b_k^{-2} \wedge a_D^{-2} \right].$$

$\square$

7.5. Proof of Proposition 5.1

Remark that, under Assumption $\mathcal{H}_{\mathcal{D}}$,

$$\frac{S_0(D)}{\varepsilon^4} \sim \sum_{\substack{k,l=1 \\ k\neq l}}^{D} k^{2t} l^{2t} |k-l|^{-2s} = 2 \sum_{k=1}^{D} k^{2t} \sum_{l=1}^{k-1} \frac{l^{2t}}{(k-l)^{2s}} \leq 2 \sum_{k=1}^{D} k^{4t} \sum_{l=1}^{k-1} \frac{1}{(k-l)^{2s}}.$$

Then, for all $k \in \mathcal{N}$,

$$\sum_{l=1}^{k-1} \frac{1}{(k-l)^{2s}} = \sum_{m=1}^{k-1} \frac{1}{m^{2s}} \sim k^{1-2s}.$$

In particular, the above sum is finite whatever the value of k provided $s > 1/2$. Hence, under Assumption $\mathcal{H}_{\mathcal{D}}$, we get

$$S_0(D) \lesssim \varepsilon^4 \sum_{k=1}^{D} k^{4t} k^{1-2s} \sim \varepsilon^4 D^{4t-2s+2} = o(R_0(D)) \quad \text{as} \quad D \to +\infty.$$

$\square$

7.6. Proof of Proposition 5.2

Recall from (7.2) that

$$\mathrm{Var}_{\theta,\xi}(T_D) = R_\theta(D) + S_\theta(D),$$

where, using (7.5),

$$R_\theta(D) = \sum_{k=1}^{D} b_k^{-4} \mathrm{Var}_{\theta,\xi}(y_k^2 - \varepsilon^2) \lesssim \varepsilon^4 \sum_{k=1}^{D} b_k^{-4} + \varepsilon^2 (\max_{1 \leq k \leq D} b_k^{-2}) \sum_{k=1}^{D} \theta_k^2.$$

Moreover, using (7.6),

$$S_\theta(D) = \sum_{\substack{k,l=1 \\ k\neq l}}^{D} b_k^{-2} b_l^{-2} \mathrm{Cov}_{\theta,\xi}(y_k^2 - \varepsilon^2, y_l^2 - \varepsilon^2)$$

$$= 2\varepsilon^4 \sum_{\substack{k,l=1 \\ k\neq l}}^{D} b_k^{-2} b_l^{-2} \mathbb{E}[(\xi_k^2 - 1)(\xi_l^2 - 1)] + 8\varepsilon^2 \sum_{\substack{k,l=1 \\ k\neq l}}^{D} b_k^{-1} b_l^{-1} \theta_k \theta_l \mathbb{E}[\xi_k \xi_l]$$

$$+ 4\varepsilon^3 \sum_{\substack{k,l=1 \\ k\neq l}}^{D} b_k^{-2} b_l^{-1} \theta_l \mathbb{E}[(\xi_k^2 - 1)\xi_l] + 4\varepsilon^3 \sum_{\substack{k,l=1 \\ k\neq l}}^{D} b_l^{-2} b_k^{-1} \theta_k \mathbb{E}[(\xi_l^2 - 1)\xi_k]$$

$$:= R_1 + R_2 + R_3 + R_4.$$

Note that, using the above proposition, $R_1 = S_0(D) = o(R_\theta(D))$ as $D \to +\infty$. Then, using (5.2), we can immediately see that $R_3 = R_4 = 0$. In order to conclude, using the Cauchy−Schwarz and Peter-Paul inequalities, we get, for any $\gamma \in]0, 1[$,

$$R_2 = 8\varepsilon^2 \sum_{\substack{k,l=1 \\ k\neq l}}^{D} b_k^{-1} b_l^{-1} \theta_k \theta_l \mathrm{Cov}(\xi_k, \xi_l) \lesssim \varepsilon^2 \sum_{\substack{k,l=1 \\ k\neq l}}^{D} b_k^{-1} b_l^{-1} |\theta_k| |\theta_l| \frac{1}{|k-l|^s}$$

$$\leq \sqrt{\varepsilon^4 \sum_{\substack{k,l=1 \\ k\neq l}}^{D} b_k^{-2} b_l^{-2} \frac{1}{|k-l|^{2s}}} \sqrt{\sum_{\substack{k,l=1 \\ k\neq l}}^{D} \theta_k^2 \theta_l^2}$$

$$\leq \gamma \Big(\sum_{k=1}^{D} \theta_k^2 \Big)^2 + \gamma^{-1} \varepsilon^4 \sum_{\substack{k,l=1 \\ k \neq l}}^{D} b_k^{-2} b_l^{-2} \frac{1}{|k-l|^{2s}} \sim \gamma \Big(\sum_{k=1}^{D} \theta_k^2 \Big)^2 + \gamma^{-1} R_1.$$

Summarizing all the above computations, we obtain, for any $\gamma \in]0,1[$,

$$\mathrm{Var}_{\theta,\xi}(T_D) \lesssim (1 + \gamma^{-1}) \varepsilon^4 \sum_{k=1}^{D} b_k^{-4} + \varepsilon^2 \big(\max_{1 \leq k \leq D} b_k^{-2} \big) \sum_{k=1}^{D} \theta_k^2 + \gamma \Big(\sum_{k=1}^{D} \theta_k^2 \Big)^2$$

$$\lesssim (1 + \gamma^{-1}) \varepsilon^4 \sum_{k=1}^{D} b_k^{-4} + \gamma \Big(\sum_{k=1}^{D} \theta_k^2 \Big)^2,$$

where we have used again the Peter-Paul inequality (see, e.g., [3], p. 18).   □

ACKNOWLEDGMENTS

REFERENCES

1. Y. Baraud, "Non-Asymptotic Minimax Rates of Testing in Signal Detection", Bernoulli **8** (5), 577−606 (2002).
2. N. Bissantz, G. Claeskens, H. Holzmann, and A. Munk, "Testing for lack of fit in inverse regression − with applications to biophotonic imaging", J. Roy. Statist. Soc., Ser. B **71** (1), 25−48 (2009).
3. R. E. Castillo and H. Rafeiro, *An Introductory Course in Lebesgue Spaces*, in *CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC* (Springer, 2016).
4. L. Cavalier, "Estimation in a Problem of Fractional Integration", Inverse Problems **20** (5), 1445−1454 (2004).
5. L. Cavalier, *Inverse Problems in Statistics*, in *Lecture Notes Statist. Proc.*, Vol. 203: *Inverse Problems and High-Dimensional Estimation* (Springer, Heidelberg, 2011), pp. 3−96.
6. Yu. I. Ingster, T. Sapatinas, and I. A. Suslina, "Minimax Nonparametric Testing in a Problem Related to the Radon Transform", Math. Methods Statist. **20** (4), 347−364 (2011).
7. Yu. I. Ingster, T. Sapatinas, and I. A. Suslina, "Minimax Signal Detection in Ill-Posed Inverse Problems", Ann. Statist. **40**, 1524−1549 (2012).
8. Yu. I. Ingster and I. A. Suslina, *Nonparametric Goodness-of-Fit Testing Under Gaussian Models*, in *Lecture Notes in Statist.* (Springer-Verlag, New York, 2003), Vol. 169.
9. L. Isserlis, "On a Formula for the Product-Moment Coefficient of any Order of a Normal Frequency Distribution in any Number of Variables", Biometrika **12** (1.2), 134−139 (1918).
10. I. M. Johnstone, "Wavelet Shrinkage for Correlated Data and Inverse Problems: Adaptivity Results", Statistica Sinica **9** (1), 51−83 (1999).
11. B. Laurent, J.-M. Loubes, and C. Marteau, "Testing Inverse Problems: A Direct or an Indirect Problem?", J. Statist. Planning and Inference **141** (5), 1849−1861 (2011).
12. B. Laurent, J.-M. Loubes, and C. Marteau, "Non-Asymptotic Minimax Rates of Testing in Signal Detection with Heterogeneous Variances", Electronic J. Statist. **6**, 91−122 (2012).
13. S. Mallat, *A Wavelet Tour of Signal Processing*, 2nd ed. (Academic Press, San Diego, 1999).
14. C. Marteau and P. Mathé, "General Regularization Schemes for Signal Detection in Inverse Problems", Math. Methods Statist. **23** (3), 176−200 (2014).
15. C. Marteau and T. Sapatinas, "A Unified Treatment for Non-Asymptotic and Asymptotic Approaches to Minimax Signal Detection", Statist. Surveys **9**, 253−297 (2015).
16. A. B. Tsybakov, *Introduction to Nonparametric Estimation*, in *Springer Series in Statist.*, Revised and extended from the 2004 French original (Springer, New York, 2009).