ISSN 1066-5307, Mathematical Methods of Statistics, 2011, Vol. 20, No. 4, pp. 347–364.

Minimax Nonparametric Testing in a Problem Related to the Radon Transform

Yu. I. Ingster[1*], T. Sapatinas[2], and I. A. Suslina[3***]**

[1]Dept. of Math. II, St. Petersburg State Electrotechn. Univ., Russia

[2]Dept. of Math. and Statist., Univ. of Cyprus, Cyprus

[3]Dept. of Math, St. Petersburg State Univ. of Inform. Technol., Mech. and Optics, Russia

Received May 24, 2011; in final form, September 5, 2011

Abstract—We consider the detection problem of a two-dimensional function from noisy observations of its integrals over lines. We study both rate and sharp asymptotics for the error probabilities in the minimax setup. By construction, the derived tests are non-adaptive. We also construct a minimax rate-optimal adaptive test of rather simple structure.

Keywords: minimax testing, Radon transform, singular value decomposition.

2000 Mathematics Subject Classification: primary 62G10, 62G20; secondary 62C20.

DOI: 10.3103/S1066530711040041

1. INTRODUCTION

The problem of tomography is to reconstruct a two-dimensional function (image) from its Radon transform, i.e., from observations of its integrals over lines. This problem, and its extension to higher dimensions, appears in different scientific fields such as radio astronomy and medical imaging (see, e.g., [7], [9], [20]). We consider the tomography problem from a statistical perspective that can be formulated as a problem of reconstructing a two-dimensional function from its noisy Radon transform (see, e.g., [5], [6], [15], [16]).

Despite some work on the minimax estimation problem of a two-dimensional function from its noisy Radon transform (see [6], [16], [17]), to the best of our knowledge, there exist no work on the corresponding minimax detection problem. The general statement of this problem is given in Section 2, while some preliminaries and notation in the minimax signal detection framework are presented in Section 3. Within this framework, in Section 4, we consider the detection problem of a two-dimensional function from its noisy Radon transform and study both rate and sharp asymptotics for the error probabilities. By construction, the derived tests are non-adaptive. A rate-optimal adaptive test of rather simple structure is also constructed. The proofs are given in the Appendix.

2. FORMULATION OF THE PROBLEM

2.1. The Radon Transform

Denote by $\|\cdot\|$ the standard Euclidean norm in $\mathbb{R}^2$, i.e., $\|x\| = (x_1^2 + x_2^2)^{1/2}$, $x = (x_1, x_2) \in \mathbb{R}^2$. Let $H = \{x \in \mathbb{R}^2 \colon \|x\| \le 1\}$ be the unit disk in $\mathbb{R}^2$, and let μ denote the Lebesgue measure in $\mathbb{R}^2$. Consider the integrals of a function $f\colon H \mapsto \mathbb{R}$ over all lines that intersect H. The lines are parameterized by the

*E-mail: yurii_ingster@mail.ru
**E-mail: t.sapatinas@ucy.ac.cy
***E-mail: isuslina@mail.ru

length $u \in [0,1]$ of the perpendicular from the origin to the line and by the orientation $\varphi \in [0, 2\pi)$ of this perpendicular. Suppose that $f \in L^1(H,\mu) \cap L^2(H,\mu)$. Define the Radon transform of the function f by

$$\mathcal{R}f(u,\varphi) = \frac{\pi}{2\sqrt{1-u^2}} \int_{-\sqrt{1-u^2}}^{\sqrt{1-u^2}} f(u\cos\varphi - t\sin\varphi, u\sin\varphi + t\cos\varphi)\,dt, \quad (u,\varphi) \in S, \tag{2.1}$$

where

$$S = \big\{(u,\varphi)\colon\, u \in [0,1],\ \varphi \in [0,2\pi)\big\}.$$

Thus, the Radon transform $\mathcal{R}f$ is π times the average of f over the line segment (parametrized by (u,φ)) that intersects H. It is natural to consider $\mathcal{R}f$ as an element of $L^2(S,\mu_0)$, where μ_0 is the measure on S defined by

$$d\mu_0(u,\varphi) = \frac{2\sqrt{1-u^2}}{\pi}\,d\varphi, \quad (u,\varphi) \in S.$$

2.2. The Gaussian White Noise Model

Consider now the Gaussian white noise model

$$dY_\varepsilon(u,\varphi) = \mathcal{R}f(u,\varphi)\,du\,d\varphi + \varepsilon\,dW(u,\varphi), \quad (u,\varphi) \in S, \tag{2.2}$$

where W is a standard Wiener sheet on S (i.e., the primitive of white noise on S) and $\varepsilon > 0$ is a small parameter (the noise level). Although this model is continuous and real data are typically discretely sampled, its versions have been extensively studied in the nonparametric literature and are considered as idealized models that provide, subject to some limitations, approximations to many sampled-data nonparametric models (see, e.g., [2], [4], [8], [21]).

The Gaussian white noise model (2.2) may also seem initially rather remote. One may, however, be helped by the observation that what it really means is the following: for any function $g \in L^2(S,\mu_0)$, the integral

$$\iint_S g(u,\varphi)\mathcal{R}f(u,\varphi)\,du\,d\varphi$$

can be observed with Gaussian error having zero mean and variance equal to $\varepsilon^2 \iint_S g^2(u,\varphi)\,du\,d\varphi$ (see, e.g., [4]).

The Radon transform $\mathcal{R}$ is a compact operator and its singular value decomposition (SVD) is well known (see, e.g., [20]). To introduce it, let $\mathbb{N} = \{1,2,\dots\}$ be the set of positive integers, put $\mathbb{Z}_+ = \mathbb{N} \cup \{0\}$, and define a set of double indices giving rise to the following lattice quadrant

$$\Gamma = \{\nu\colon \nu = (j,l),\ j,l \in \mathbb{Z}_+\}. \tag{2.3}$$

An orthonormal complex-valued basis for $L^2(H,\mu)$ is given by

$$\tilde{\phi}_\nu(r,\theta) = \pi^{-1/2}(j+l+1)^{1/2} Z_{j+l}^{|j-l|}(r)\exp\{i(j-l)\theta\}, \quad \nu \in \Gamma, \tag{2.4}$$

where $x = (r\cos\theta, r\sin\theta) \in H$, with Z_a^b denoting the Zernike polynomial of degree a and order b, with $a,b \in \mathbb{Z}_+$ (see, e.g., [7]). The corresponding orthonormal complex-valued basis in $L^2(S,\mu_0)$ is

$$\tilde{\psi}_\nu(u,\varphi) = \pi^{-1/2}U_{j+l}(u)\exp\{i(j-l)\varphi\}, \quad \nu \in \Gamma, \quad (u,\varphi) \in S, \tag{2.5}$$

where

$$U_m(\cos\theta) = \frac{\sin((m+1)\theta)}{\sin\theta}, \quad m \in \mathbb{Z}_+, \quad \theta \in [0,2\pi),$$

are the Chebyshev polynomials of the second kind. We then have (see, e.g., [6])

$$\mathcal{R}\tilde{\phi}_\nu = b_\nu\tilde{\psi}_\nu$$

with singular values

$$b_\nu = \pi(j+l+1)^{-1/2}, \quad \nu \in \Gamma. \tag{2.6}$$

Since we work with real-valued functions f, the complex-valued bases (2.4) and (2.5) are identified, in standard fashion, with the equivalent real-valued orthonormal bases ϕ_ν and ψ_ν, $\nu \in \Gamma$, respectively, defined by

$$\phi_\nu = \begin{cases} \sqrt{2}\,\mathrm{Re}(\tilde{\phi}_\nu) & \text{if } j > l, \\ \tilde{\phi}_\nu & \text{if } j = l, \\ \sqrt{2}\,\mathrm{Im}(\tilde{\phi}_\nu) & \text{if } j < l, \end{cases} \tag{2.7}$$

with an analogous expression for ψ_ν, $\nu \in \Gamma$.

Hence, by standard calculations (see, e.g., [6], [13]) and an application of the spectral theorem for the self-adjoint compact operator $\mathcal{R}^*\mathcal{R}$ ($\mathcal{R}^*$ being the adjoint of $\mathcal{R}$), the Gaussian white noise model (2.2) generates the following equivalent discrete observational model in the Fourier domain, called the Gaussian sequence model,

$$y_\nu = b_\nu \theta_\nu + \varepsilon\, \xi_\nu, \quad \nu \in \Gamma, \tag{2.8}$$

where $y_\nu = \langle \mathcal{R}f, \psi_\nu \rangle$, $\nu \in \Gamma$, are the “observations”, b_ν, $\nu \in \Gamma$, are the singular values of the Radon operator $\mathcal{R}$ given by (2.6), $\theta_\nu = \langle f, \phi_\nu \rangle$, $\nu \in \Gamma$, are the Fourier coefficients of f with respect to ϕ_ν given by (2.7), and ξ_ν, $\nu \in \Gamma$, are independent and identically distributed (iid) standard Gaussian random variables, i.e., $\xi_\nu \overset{iid}{\sim} \mathcal{N}(0,1)$, $\nu \in \Gamma$.

2.3. The Class of Functions

Crucial to the suggested detection methodology is the idea of considering minimax detection over certain classes of functions in $f \in L^2(H,\mu)$. Following [6], we consider a special class of functions with polynomially decreasing coefficients $\theta = \{\theta_\nu\}_{\nu \in \Gamma}$, i.e., for some $p > 0$, $L > 0$,

$$\mathcal{F}(p,L) = \Big\{ f = \sum_{\nu \in \Gamma} \theta_\nu \phi_\nu \colon \theta \in \tilde{\Theta}(p,L) \Big\} \tag{2.9}$$

with

$$\tilde{\Theta}(p,L) = \Big\{ \theta \in l^2 \colon \sum_{\nu \in \Gamma,\, \nu \neq (0,0)} (j+1)^{2p}(l+1)^{2p}\theta_\nu^2 \le L^2 \Big\}. \tag{2.10}$$

It has been shown that $\mathcal{F}(p,L)$ can be identified with the set of functions f which have $2p$ weak derivatives (provided $2p$ is an integer) that are square-integrable on H with respect to the modified dominating measure

$$d\mu_{2p+1}(x) = (1 - \|x\|^2)^{2p}\, d\mu(x), \quad x \in H.$$

This is weaker than the square-integrability with respect to μ assumed for the usual Sobolev spaces (see Proposition 2.2 in [14]).

2.4. The Aim

The goal is to determine whether the two-dimensional function f corresponds to a known “etalon" function f_0 (i.e., to test the null hypothesis $H_0\colon f = f_0$) or there exists a difference between f and f_0 (i.e., against the alternative hypothesis $H_1\colon f = f_0 + \Delta f$ with $\Delta f \in \mathcal{F}(p,L)$, see (2.9)–(2.10)), based on the observation of a trajectory $\{Y_\varepsilon = Y_\varepsilon(u,\varphi)\}$, $(u,\varphi) \in S$, from the Guassian white noise model (2.2).

From mathematical point of view, we can take $f_0 = 0$ by passing to the observation $\tilde{Y}_\varepsilon$ with $d\tilde{Y}_\varepsilon = dY_\varepsilon(u,\phi) - \mathcal{R}f_0\, du\, d\phi$. For this reason, without loss of generality, we assume in the sequel that $f_0 = 0$, use f in place of Δf, and take the observation Y_ε. In order to avoid having a trivial power (see below), our ultimate goal is to determine whether f satisfies (3.1) (see below), using only tests calibrated in such a way that if one had run them in the absence of an $f \in \mathcal{F}(p, L)$, a certain restriction of the significance level (error probability) is met.

In the sequel, we elaborate on the set under the alternative hypothesis and the suggested test statistics that provide a good quality of testing in the minimax framework. Before going into the details, however, we give the necessary preliminaries on the minimax signal detection framework in the standard Gaussian white noise model which provide the avenue for developing the suggested detection methodology and deriving theoretical results for detecting a two-dimensional function from its noisy Radon transform.

Hereafter, the relation $A_\varepsilon \sim B_\varepsilon$ means that $A_\varepsilon/B_\varepsilon \to 1$ as $\varepsilon \to 0$ while the relation $A_\varepsilon \asymp B_\varepsilon$ means that there exist absolute constants $0 < c_1 \le c_2 < \infty$ and $\varepsilon_0 > 0$ small enough such that $c_1 \le A_\varepsilon/B_\varepsilon \le c_2$ for $0 < \varepsilon \le \varepsilon_0$.

3. SIGNAL DETECTION IN THE GAUSSIAN SEQUENCE MODEL: THE MINIMAX FRAMEWORK

Consider the Gaussian sequence model (2.8). In order to avoid having a trivial minimax hypothesis testing problem (i.e., trivial power), one usually needs to remove a neighborhood around the functional parameter under the null hypothesis and to impose some additional constraints that are typically expressed in the form of some regularity conditions, such as constraints on the derivatives of the unknown functional parameter of interest (see, e.g., [11], Sections 1.3–1.4).

In view of the above observation, the main object of our study is the hypothesis testing problem

$$H_0\colon \theta = 0 \quad \text{versus} \quad H_1\colon \sum_{\nu\in\Gamma} a_\nu^2\theta_\nu^2 \le 1, \quad \sum_{\nu\in\Gamma}\theta_\nu^2 \ge r_\varepsilon^2, \tag{3.1}$$

where $\theta = \{\theta_\nu\}_{\nu\in\Gamma} \in l^2$, $a_\nu \ge 0$, $\nu \in \Gamma$, and $r_\varepsilon > 0$, $r_\varepsilon \to 0$, is a given family. It means that the alternative set corresponds to an ellipsoid of semi-axes $1/a_\nu$, $\nu \in \Gamma$, with an l^2-ball of radius r_ε removed. (Here, $l^2 = \{\zeta\colon \sum_{\nu\in\Gamma}\zeta_\nu^2 < \infty\}$ with Γ given by (2.3).)

Consider now the sequence $\eta = \{\eta_\nu\}_{\nu\in\Gamma}$ with elements $\eta_\nu = \theta_\nu/\sigma_\nu$, where we set $\sigma_\nu = 1/b_\nu$, $\nu \in \Gamma$. In view of (2.6), the sequence $\eta = \{\eta_\nu\}_{\nu\in\Gamma} \in l_2$, and the Gaussian sequence model (2.8) takes the form

$$y_\nu = \eta_\nu + \varepsilon\xi_\nu, \quad \nu \in \Gamma. \tag{3.2}$$

The hypothesis testing problem (3.1) can now be written in the following equivalent form

$$H_0\colon \eta = 0 \quad \text{versus} \quad H_1\colon \eta \in \Theta(r_\varepsilon), \tag{3.3}$$

where the alternative set, i.e., $\Theta(r_\varepsilon)$, is determined by the constraints

$$\Theta = \Big\{\eta \in l^2\colon \sum_{\nu\in\Gamma} a_\nu^2\sigma_\nu^2\eta_\nu^2 \le 1\Big\}, \qquad \Theta(r_\varepsilon) = \Big\{\eta \in \Theta\colon \sum_{\nu\in\Gamma}\sigma_\nu^2\eta_\nu^2 \ge r_\varepsilon^2\Big\}, \tag{3.4}$$

i.e., the alternative set corresponds to an ellipsoid of semi-axes $1/(a_\nu\sigma_\nu)$, $\nu \in \Gamma$, with an ellipsoid of semi-axes r_ε/σ_ν, $\nu \in \Gamma$, removed.

We are therefore interested in the minimax efficiency of the hypothesis testing problem (3.3)–(3.4) for a given family of sets $\Theta_\varepsilon = \Theta(r_\varepsilon) \subset l^2$. It is characterized by asymptotics, as $\varepsilon \to 0$, of the minimax error probabilities in the problem at hand. Namely, for a (randomized) test ψ (i.e., a measurable function of the observation $y = \{y_\nu\}_{\nu\in\Gamma}$ taking values in $[0, 1]$), the null hypothesis is rejected with probability $\psi(y)$ and is accepted with probability $1 - \psi(y)$. Let $P_{\varepsilon,\eta}$ be the probability measure for the Gaussian sequence model (3.2) and denote by $E_{\varepsilon,\eta}$ the expectation over this probability measure. Let $\alpha_\varepsilon(\psi) = E_{\varepsilon,0}\psi$ be its type I error probability, and let $\beta_\varepsilon(\Theta_\varepsilon, \psi) = \sup_{\eta\in\Theta_\varepsilon} E_{\varepsilon,\eta}(1 - \psi)$ be its maximal type II error probability. We consider two criteria of asymptotic optimality:

(1) The first one corresponds to the classical Neyman–Pearson criterion. For $\alpha \in (0, 1)$, we set

$$\beta_\varepsilon(\Theta_\varepsilon, \alpha) = \inf_{\psi\colon \alpha_\varepsilon(\psi)\le\alpha} \beta_\varepsilon(\Theta_\varepsilon, \psi).$$

We call a family of tests $\psi_{\varepsilon,\alpha}$ *asymptotically minimax* if

$$\alpha_\varepsilon(\psi_{\varepsilon,\alpha}) \le \alpha + o(1), \qquad \beta_\varepsilon(\Theta_\varepsilon, \psi_{\varepsilon,\alpha}) = \beta_\varepsilon(\Theta_\varepsilon, \alpha) + o(1),$$

where $o(1)$ is a family tending to zero; here and in what follows, all limits are taken as $\varepsilon \to 0$ unless otherwise stated.

(2) The second one corresponds to the total error probabilities. Let $\gamma_\varepsilon(\Theta_\varepsilon, \psi)$ be the sum of the type I and the maximal type II error probabilities, and let $\gamma_\varepsilon(\Theta_\varepsilon)$ be the minimax total error probability, i.e.,

$$\gamma_\varepsilon(\Theta_\varepsilon) = \inf_\psi \gamma_\varepsilon(\Theta_\varepsilon, \psi),$$

where the infimum is taken over all possible tests. We call a family of tests ψ_ε *asymptotically minimax* if

$$\gamma_\varepsilon(\Theta_\varepsilon, \psi_\varepsilon) = \gamma_\varepsilon(\Theta_\varepsilon) + o(1).$$

It is known that (see, e.g., [11], Chapter 2)

$$\beta_\varepsilon(\Theta_\varepsilon, \alpha) \in [0, 1-\alpha], \qquad \gamma_\varepsilon(\Theta_\varepsilon) = \inf_{\alpha\in(0,1)} (\alpha + \beta_\varepsilon(\Theta_\varepsilon, \alpha)) \in [0, 1]. \tag{3.5}$$

We consider the problems of rate and sharp asymptotics for the error probabilities in the minimax setup. The rate optimality problem corresponds to the study of the conditions for which $\gamma_\varepsilon(\Theta_\varepsilon) \to 1$ or $\gamma_\varepsilon(\Theta_\varepsilon) \to 0$ and, in the latter case, to the construction of *asymptotically minimax consistent* families of tests ψ_ε, i.e., such that $\gamma_\varepsilon(\Theta_\varepsilon, \psi_\varepsilon) \to 0$.

We are interested in a set Θ_ε of the form

$$\Theta_\varepsilon = \Theta(r_\varepsilon) = \{\eta \in \Theta \colon |\eta| \geq r_\varepsilon\},$$

where $\Theta \subset l_2$ is a given set, $|\cdot|$ is some norm in l_2 (not necessarily the standard l_2-norm) and $r_\varepsilon \to 0$ is a given positive-valued family. For this case, we use the notation $\gamma_\varepsilon(\Theta(r_\varepsilon)) = \gamma_\varepsilon(r_\varepsilon)$, $\beta_\varepsilon(\Theta(r_\varepsilon), \alpha) = \beta_\varepsilon(r_\varepsilon, \alpha)$ and we are interested in the minimal decreasing rates for the sequence r_ε such that $\gamma_\varepsilon(r_\varepsilon) \to 0$. Namely, we say that a positive sequence $r^*_\varepsilon \to 0$ is a *separation rate*, if

$$\gamma_\varepsilon(r_\varepsilon) \to 1 \quad \text{and} \quad \beta_\varepsilon(r_\varepsilon, \alpha) \to 1-\alpha \quad \text{for any} \quad \alpha \in (0,1) \quad \text{as} \quad r_\varepsilon/r^*_\varepsilon \to 0, \tag{3.6}$$

and

$$\gamma_\varepsilon(r_\varepsilon) \to 0 \quad \text{and} \quad \beta_\varepsilon(r_\varepsilon, \alpha) \to 0 \quad \text{for any} \quad \alpha \in (0,1) \quad \text{as} \quad r_\varepsilon/r^*_\varepsilon \to \infty. \tag{3.7}$$

In other words, it means that, for small ε, one can detect all sequences $\eta \in \Theta(r_\varepsilon)$ if the ratio $r_\varepsilon/r^*_\varepsilon$ is large, whereas if this ratio is small then it is impossible to distinguish between the null and the alternative hypotheses with small minimax total error probability. Hence the rate optimality problem corresponds to finding the separation rates r^*_ε and to constructing asymptotically minimax consistent families of tests.

On the other hand, the sharp optimality problem corresponds to the study of the asymptotics of the quantities $\beta_\varepsilon(\Theta_\varepsilon, \alpha)$, $\gamma_\varepsilon(\Theta_\varepsilon)$ (up to vanishing terms) and to the construction of asymptotically minimax families of tests $\psi_{\varepsilon,\alpha}$ and ψ_ε, respectively. Often, the sharp asymptotics are of Gaussian type, i.e.,

$$\beta_\varepsilon(\Theta_\varepsilon, \alpha) = \Phi(H^{(\alpha)} - u_\varepsilon) + o(1), \qquad \gamma_\varepsilon(\Theta_\varepsilon) = 2\Phi(-u_\varepsilon/2) + o(1), \tag{3.8}$$

where Φ is the standard Gaussian distribution function, $H^{(\alpha)}$ is its $(1-\alpha)$-quantile, i.e., $\Phi(H^{(\alpha)}) = 1-\alpha$. The quantity $u_\varepsilon = u_\varepsilon(r_\varepsilon)$ is the value of the specific extreme problem (4.1) on the sequence space l^2, and the extreme sequence of this problem determines the structure of the asymptotically minimax families of tests $\psi_{\varepsilon,\alpha}$ and ψ_ε. Moreover, we shall see that if $u_\varepsilon(r_\varepsilon) \to \infty$, then $\gamma_\varepsilon(r_\varepsilon) \to 0$, $\beta_\varepsilon(r_\varepsilon, \alpha) \to 0$, and if $u_\varepsilon(r_\varepsilon) \to 0$, then $\gamma_\varepsilon(r_\varepsilon) \to 1$, $\beta_\varepsilon(r_\varepsilon, \alpha) \to 1-\alpha$ for any $\alpha \in (0,1)$, i.e., the family $u_\varepsilon(r_\varepsilon)$ characterizes *distinguishability* in the testing problem. The separation rates r^*_ε are usually determined by the relation $u_\varepsilon(r^*_\varepsilon) \asymp 1$ (see, e.g., [10], [11]). Hence sharp and rate optimality problems correspond to the study of the extreme problem (4.1) and of the asymptotics of the family $u_\varepsilon(r_\varepsilon)$.

4. MINIMAX IMAGE DETECTION FROM NOISY TOMOGRAPHIC DATA

4.1. A General Result: Rate and Sharp Asymptotics

Recall the Gaussian sequence model (3.2). We are interested in the hypothesis testing problem (3.3) with the alternative set $\Theta_\varepsilon = \Theta(r_\varepsilon)$ given by (3.4).

Consider now the extreme problem

$$u_\varepsilon^2 = u_\varepsilon^2(r_\varepsilon) = \frac{1}{2\varepsilon^4} \inf_{\eta \in \Theta(r_\varepsilon)} \sum_{\nu \in \Gamma} \eta_\nu^4. \tag{4.1}$$

Suppose that $\Theta(r_\varepsilon) \neq \emptyset$ and $u_\varepsilon > 0$, and let there exist an extreme sequence $\{\tilde{\eta}_\nu\}_{\nu \in \Gamma}$ in the extreme problem (4.1). (Observe the uniqueness of a nonnegative extreme sequence $\{\tilde{\eta}_\nu\}_{\nu \in \Gamma}$, because, by passing to the sequence $\{z_\nu\}_{\nu \in \Gamma}$ with elements $z_\nu = \tilde{\eta}_\nu^2$, $\nu \in \Gamma$, we obtain the minimization problem of a strictly convex function under linear constraints.) Denote

$$w_\nu = \frac{\tilde{\eta}_\nu^2}{\sqrt{2 \sum_{\nu \in \Gamma} \tilde{\eta}_\nu^4}}, \quad \nu \in \Gamma, \qquad w_0 = \sup_{\nu \in \Gamma} w_\nu, \tag{4.2}$$

and consider the following families of test statistics and tests

$$t_\varepsilon = \sum_{\nu \in \Gamma} w_\nu((y_\nu/\varepsilon)^2 - 1), \qquad \psi_{\varepsilon,H} = \mathbf{1}_{\{t_\varepsilon > H\}}, \tag{4.3}$$

where $\mathbf{1}_{\{A\}}$ denotes the indicator function of a set A. (Note that the values of $\tilde{\eta}_\nu$, w_ν, $\nu \in \Gamma$, and w_0 depend on ε, i.e., $\tilde{\eta}_\nu = \tilde{\eta}_{\nu,\varepsilon}$, $w_\nu = w_{\nu,\varepsilon}$, $\nu \in \Gamma$, and $w_0 = w_{0,\varepsilon}$.)

The key tool for the study of the above mentioned hypothesis testing problem is the following general theorem. Its proof follows along the lines of the proof of Theorem 4.1 in [13]; hence it is omitted.

Theorem 4.1. *Consider the Gaussian sequence model* (3.2) *and the hypothesis testing problem* (3.3) *with the alternative set given by* (3.4). *Let u_ε be determined by the extreme problem* (4.1), *let the coefficients w_ν, $\nu \in \Gamma$, and w_0 be as in* (4.2), *and consider the family of tests $\psi_{\varepsilon,H}$ given by* (4.3). *Then*

(1) (a) *If $u_\varepsilon \to 0$, then $\beta_\varepsilon(r_\varepsilon, \alpha) \to 1 - \alpha$ for any $\alpha \in (0,1)$ and $\gamma_\varepsilon(r_\varepsilon) \to 1$, i.e., minimax testing is impossible. If $u_\varepsilon = O(1)$, then* $\liminf \beta_\varepsilon(r_\varepsilon, \alpha) > 0$ *for any $\alpha \in (0,1)$ and* $\liminf \gamma_\varepsilon(r_\varepsilon) > 0$, *i.e., minimax consistent testing is impossible.*

(b) *If $u_\varepsilon \asymp 1$ and $w_0 = o(1)$, then the tests $\psi_{\varepsilon,H}$ of the form* (4.3) *with $H = H^{(\alpha)}$ and $H = u_\varepsilon/2$ are asymptotically minimax, i.e.,*

$$\alpha_\varepsilon(\psi_{\varepsilon,H^{(\alpha)}}) \le \alpha + o(1),$$
$$\beta_\varepsilon(\Theta(r_\varepsilon), \psi_{\varepsilon,H^{(\alpha)}}) = \beta_\varepsilon(r_\varepsilon, \alpha) + o(1),$$
$$\gamma_\varepsilon(\Theta(r_\varepsilon), \psi_{\varepsilon,u_\varepsilon/2}) = \gamma_\varepsilon(r_\varepsilon) + o(1),$$

and the sharp asymptotics (3.8) *hold true, i.e.,*

$$\beta_\varepsilon(r_\varepsilon, \alpha) = \Phi(H^{(\alpha)} - u_\varepsilon) + o(1),$$
$$\gamma_\varepsilon(r_\varepsilon) = 2\Phi(-u_\varepsilon/2) + o(1).$$

(2) *If $u_\varepsilon \to \infty$, then the tests $\psi_{\varepsilon,H}$ of the form* (4.3) *with $H = T_\varepsilon$ are asymptotically minimax consistent for any $c \in (0,1)$ and a family $T_\varepsilon \sim cu_\varepsilon$, i.e., $\gamma_\varepsilon(\Theta(r_\varepsilon), \psi_{\varepsilon,T_\varepsilon}) \to 0$.*

Theorem 4.1 shows that the asymptotics of the quality of testing is determined by the asymptotics of values u_ε of the extreme problem (4.1). In order to make use of it, one needs to study the extreme problem (4.1). This problem is studied by using Lagrange multipliers. Then, the extreme sequence in the above mentioned extreme problem is of the form

$$\tilde{\eta}_\nu^2 = z_0^2 \sigma_\nu^2 (1 - A a_\nu^2)_+, \quad \nu \in \Gamma, \tag{4.4}$$

where $(t)_+ = \max(t, 0)$, $t \in \mathbb{R}$, and the quantities $z_0 = z_{0,\varepsilon}$ and $A = A_\varepsilon$ are determined by the equations

$$\begin{cases} \sum_{\nu\in\Gamma} \sigma_\nu^2 \tilde{\eta}_\nu^2 = r_\varepsilon^2, \\ \sum_{\nu\in\Gamma} a_\nu^2 \sigma_\nu^2 \tilde{\eta}_\nu^2 = 1. \end{cases} \tag{4.5}$$

The equations (4.5) are immediately rewritten in the form

$$\begin{cases} r_\varepsilon^2 = z_0^2 J_1, \\ 1 = z_0^2 A^{-1} J_2, \end{cases} \tag{4.6}$$

and, hence, the extreme problem (4.1) takes the form

$$u_\varepsilon^2 = \varepsilon^{-4} z_0^4 J_0/2, \tag{4.7}$$

where

$$J_1 = \sum_{\nu\in\Gamma} \sigma_\nu^4 (1 - A a_\nu^2)_+, \qquad J_2 = A \sum_{\nu\in\Gamma} a_\nu^2 \sigma_\nu^4 (1 - A a_\nu^2)_+, \qquad J_0 = J_1 - J_2 = \sum_{\nu\in\Gamma} \sigma_\nu^4 (1 - A a_\nu^2)_+^2.$$

It is also convenient to rewrite (4.6) and (4.7) in the form

$$r_\varepsilon^2 = A\,\frac{J_1}{J_2}, \qquad u_\varepsilon^2 = \left(\frac{r_\varepsilon}{\varepsilon}\right)^4 \frac{J_0}{2J_1^2}. \tag{4.8}$$

Remark 4.1. Let $u_\varepsilon = u_\varepsilon(r_\varepsilon)$ be the value of the extreme problem (4.1) with sequences $a = \{a_\nu\}_{\nu\in\Gamma}$ and $\sigma = \{\sigma_\nu\}_{\nu\in\Gamma}$ associated with the alternative set $\Theta_\varepsilon = \Theta(r_\varepsilon)$ given by (3.4), and let $\tilde{u}_\varepsilon = \tilde{u}_\varepsilon(r_\varepsilon)$ be the corresponding value of the extreme problem similar to (4.1) with sequences $\tilde{a} = Ca = \{Ca_\nu\}_{\nu\in\Gamma}$ and $\tilde{\sigma} = D\sigma = \{D\sigma_\nu\}_{\nu\in\Gamma}$ in (3.4), for some constants $C, D > 0$. Then, it is easily seen that the relation $\tilde{u}_\varepsilon(r_\varepsilon) = (CD)^{-2} u_\varepsilon(C r_\varepsilon)$ holds true.

Remark 4.2. In order to obtain the corresponding rate and sharp asymptotics for the noisy tomographic data, we need to study the asymptotics of the quantities J_i, $i = 0, 1, 2$, given above. We note, however, that the methods used in [13] to study analogous asymptotics in a wide range of linear statistical ill-posed inverse problems cannot be adapted to the problem at hand. The reason is that there does not exist a common ordering for the sequences $a = \{a_\nu\}_{\nu\in\Gamma}$ and $\sigma = \{\sigma_\nu\}_{\nu\in\Gamma}$ associated with the alternative set $\Theta_\varepsilon = \Theta(r_\varepsilon)$ given by (3.4). The arguments and techniques used to prove Theorems 4.2 and 4.3 below are specifically developed to tackle this problem.

4.2. Rate and Sharp Asymptotics for the Noisy Tomographic Data

According to (2.6) and (2.10), consider the double-index sequences

$$a_\nu = L^{-1}(j+1)^p (l+1)^p, \quad \nu \in \Gamma, \tag{4.9}$$

$$\sigma_\nu = \pi^{-1}(j+l+1)^{1/2}, \quad \nu \in \Gamma, \tag{4.10}$$

for some $p > 0$ with Γ given by (2.3).

Theorem 4.2. *Consider the Gaussian sequence model* (3.2) *and the hypothesis testing problem* (3.3) *with the alternative set given by* (3.4). *Let* $\{a_\nu\}_{\nu\in\Gamma}$ *and* $\{\sigma_\nu\}_{\nu\in\Gamma}$ *be defined as in* (4.9) *and* (4.10), *respectively. Then*

(a) *The sharp asymptotics* (3.8) *hold with the value* u_ε *of the extreme problem* (4.1) *determined by*

$$u_\varepsilon^2 \sim \pi^4 L^{-3/p} r_\varepsilon^{4+3/p} \varepsilon^{-4} \frac{2p+3}{2B} \left(\frac{3}{4p+3}\right)^{1+3/(2p)}, \tag{4.11}$$

where $B = \sum_{m=1}^\infty m^{-3} = \zeta(3) \approx 1.202$ *with* $\zeta(\cdot)$ *the Riemann's zeta-function.*

(b) *The asymptotically minimax family of tests* $\psi_{\varepsilon,H}$ *are determined by the family of test statistics* t_ε *given by* (4.3) *with coefficients* w_ν, $\nu \in \Gamma$, *and* w_0 *as in* (4.2), *and with extreme sequence* $\{\tilde{\eta}_\nu\}_{\nu\in\Gamma}$ *satisfying* (4.4) *with* $A \sim \frac{3}{4p+3} r_\varepsilon^2$.

(c) *The separation rates are of the form*

$$r_\varepsilon^* = \varepsilon^{4p/(4p+3)}. \tag{4.12}$$

Remark 4.3. It is easily seen that the asymptotic results in Theorem 4.2 hold true uniformly over $p \in \Sigma$ for any compact set $\Sigma \subset (0, \infty)$.

Remark 4.4. Rate and sharp asymptotics in the corresponding minimax estimation problem under the L^2-risk have been obtained in [6]. In particular, the asymptotic (as $\varepsilon \to 0$) minimax rates of estimation are given by

$$R_\varepsilon^2 := \inf_{\tilde{f}} \sup_{f\in\mathcal{F}(p,L)} E||\tilde{f} - f||^2 \asymp \varepsilon^{4p/(2p+2)},$$

where the infimum is taken over all possible estimators $\tilde{f}$ of f based on observations from the Gaussian white noise model (2.2). (Here, we adopt the standard notation and write $g_1(\varepsilon) \asymp g_2(\varepsilon)$ to denote $0 < \liminf(g_1(\varepsilon)/g_2(\varepsilon)) \le \limsup(g_1(\varepsilon)/g_2(\varepsilon)) < \infty$ as $\varepsilon \to 0$.) By comparing r_ε^* with R_ε, it is observed that the asymptotic minimax rates of testing are faster than the corresponding asymptotic minimax rates of estimation; this phenomenon is common in nonparametric statistical inference (see, e.g., [11], Sections 2.10 and 3.5.1, [13]).

4.3. Adaptivity and Rate Optimality for the Noisy Tomographic Data

The family of tests considered in Section 4.2 depends on the parameter p that is usually unknown in practice. Therefore, it is of paramount importance to construct families of tests that do not depend on the unknown parameter p and, at the same time, provide the best possible asymptotic minimax efficiency. Such families are called *adaptive* (to the parameter p), and the formal setting is as follows.

Let Σ be a compact set in $(0, \infty)$ and a family $r_\varepsilon(p)$, $p \in \Sigma$, be given, where $\varepsilon > 0$ is small. Let the set $\Theta_\varepsilon(p, r_\varepsilon(p))$ be determined by the constraints (3.4) with $a_\nu = a_\nu(p)$, $\nu \in \Gamma$, and $r_\varepsilon = r_\varepsilon(p)$, and set

$$\Theta_\varepsilon(\Sigma) = \bigcup_{p\in\Sigma} \Theta_\varepsilon(p, r_\varepsilon(p)).$$

We are interested in the following hypothesis testing problem

$$H_0\colon \eta = 0, \quad \text{versus} \qquad H_1\colon \eta \in \Theta_\varepsilon(\Sigma).$$

We aim at finding conditions for either $\gamma_\varepsilon(\Theta_\varepsilon(\Sigma)) \to 1$ or $\gamma_\varepsilon(\Theta_\varepsilon(\Sigma)) \to 0$, and to constructing asymptotically minimax adaptive consistent families of tests ψ_ε^{ad} such that $\gamma_\varepsilon(\Theta_\varepsilon(\Sigma), \psi_\varepsilon^{ad}) \to 0$ as $\gamma_\varepsilon(\Theta_\varepsilon(\Sigma)) \to 0$.

Let $u_\varepsilon(p) = u_\varepsilon(p, r_\varepsilon(p))$ be the value of the extreme problem (4.1) for the set $\Theta_\varepsilon = \Theta_\varepsilon(p, r_\varepsilon(p))$. Put

$$u_\varepsilon(\Sigma) = \inf_{p\in\Sigma} u_\varepsilon(p).$$

We are interested in how large $u_\varepsilon(\Sigma)$ should be in order to provide the relation $\gamma_\varepsilon(\Theta_\varepsilon(\Sigma)) \to 0$. We say that the family $u_\varepsilon^{ad} = u_\varepsilon^{ad}(\Sigma) \to \infty$ characterizes *adaptive distinguishability* if there exist constants $0 < d = d(\Sigma) \le D(\Sigma) = D < \infty$ such that

$$\gamma_\varepsilon(\Theta_\varepsilon(\Sigma)) \to 1 \quad \text{as} \quad \limsup_{p\in\Sigma} u_\varepsilon(p)/u_\varepsilon^{ad} < d,$$

$$\gamma_\varepsilon(\Theta_\varepsilon(\Sigma)) \to 0 \quad \text{as} \quad \liminf_{p\in\Sigma} u_\varepsilon(p)/u_\varepsilon^{ad} > D.$$

Observe that it follows from the asymptotics (4.11) that, by making $r_\varepsilon(p)$ larger or smaller, one can increase or decrease $u_\varepsilon(p, r_\varepsilon(p))$ in order to get $u_\varepsilon(p, r_\varepsilon(p)) \sim u_\varepsilon$, for all $p \in \Sigma$ and any family $u_\varepsilon > 0$.

We call a family $r_\varepsilon^{ad}(p),\ p \in \Sigma$, such that $u_\varepsilon^{ad} \asymp u_\varepsilon(p, r_\varepsilon^{ad}(p))$, the family of *adaptive separation rates*.

Note that the relation $\gamma_\varepsilon(\Theta_\varepsilon(\Sigma)) \to 0$ is possible if $u_\varepsilon(\Sigma) \to \infty$. However this implication does not hold for the tomography problem under consideration, as we show below. (A similar situation appears in some ill-posed inverse problems, see [13].) Hence, hereafter, adaptive distinguishability conditions and adaptive separation rates are sought for the tomography problem. In contrast to Theorem 4.2, there is price to pay for the adaptation. We show below that

$$u_\varepsilon^{ad} = \sqrt{\log\log \varepsilon^{-1}}, \tag{4.13}$$

yielding a loss in the separation rates in terms of an extra factor $\sqrt[4]{\log\log\varepsilon^{-1}}$ in ε. Furthermore, the derived families of tests are of simple structure. (A similar loss in the separation rates was first observed in [22] and more recently in [13].)

Specifically, let p be unknown, $p \in \Sigma$, where $\Sigma = [p_{\min}, p_{\max}]$, and let $0 < p_{\min} < p_{\max} < \infty$ be a compact interval in $(0, \infty)$. Let us also consider the collections

$$p_k \in \Sigma, \quad c_k \sim 2/r_\varepsilon(p_k), \quad k = 0, 1, \dots, K, \qquad K = K_\varepsilon \asymp \log(\varepsilon^{-1}) \log\log(\varepsilon^{-1}),$$

where $p_0 = p_{\max} > p_1 > \dots > p_K = p_{\min}$ (the collection $p_k,\ k = 1, 2, \dots, K-1$, will be specified in the proof) and collection of statistics t_{ε, c_k} of the form

$$t_{\varepsilon,c_k} = \sum_{\nu \in C_{\nu,k}} w_{\nu,k}\big((y_\nu^2/\varepsilon^2) - 1\big), \qquad w_{\nu,k} = \frac{\sigma_\nu^2}{\big(2\sum_{\nu\in C_{\nu,k}} \sigma_\nu^4\big)^{1/2}}, \qquad \sum_{\nu\in C_{\nu,k}} w_{\nu,k}^2 = \frac{1}{2}, \tag{4.14}$$

for $\nu \in \Gamma$ given by (2.3) and $C_{\nu,k} = \{\nu\colon a_{\nu,p_k} \le c_k\}$. Consider the following families of thresholds and tests

$$H_\varepsilon = 2\sqrt{\log(K_\varepsilon)}, \qquad \mathcal{Y}_\varepsilon = \{y\colon t_{\varepsilon,c_k} \le H_\varepsilon,\ \forall\, 0 \le k \le K_\varepsilon\}, \qquad \psi_\varepsilon = \mathbf{1}_{\overline{\mathcal{Y}}_\varepsilon}, \tag{4.15}$$

where $\bar{A}$ denotes the complement of a set A.

Denote also

$$\phi(p) = \frac{4}{4p+3}, \qquad \phi(\Sigma) = \{\phi(p)\colon p \in \Sigma\} \subset (0, \infty). \tag{4.16}$$

Theorem 4.3. *Consider the Gaussian sequence model* (3.2) *and the hypothesis testing problem* (3.3) *with the alternative set given by* (3.4). *Let* $\{a_\nu\}_{\nu\in\Gamma}$ *and* $\{\sigma_\nu\}_{\nu\in\Gamma}$ *be defined as in* (4.9) *and* (4.10)*, respectively. Then*

(a) (*lower bounds*) *Let the set* $\phi(\Sigma)$ *given by* (4.16) *contains an interval* $[a, b],\ 0 < a < b < 4/3$. *There exists constant* $d = d(\Sigma) > 0$ *such that if* $\limsup_{p\in\Sigma} u_\varepsilon(p)/\sqrt{\log\log(\varepsilon^{-1})} \le d$*, then* $\gamma_\varepsilon(\Theta_\varepsilon(\Sigma)) \to 1$.

(b) (*upper bounds*) *For the family of tests* ψ_ε *given by* (4.15)*,* $\alpha(\psi_\varepsilon) = o(1)$ *and there exists a constant* $D = D(\Sigma) > 0$ *such that if* $\liminf_{p\in\Sigma} u_\varepsilon(p)/\sqrt{\log\log(\varepsilon^{-1})} > D$*, then* $\beta_\varepsilon(\Theta_\varepsilon(\Sigma), \psi_\varepsilon) = o(1)$.

(c) (*adaptive separation rates*) *The adaptive distinguishability family* u_ε^{ad} *is given by* (4.13) *and the adaptive separation rates* $r_\varepsilon^{ad}(p)$*,* $p \in \Sigma$*, are given by*

$$r_\varepsilon^{ad}(p) = \big(\varepsilon \sqrt[4]{\log\log(\varepsilon^{-1})}\,\big)^{4p/(4p+3)}.$$

Remark 4.5. Rate and sharp adaptation in the corresponding minimax estimation problem under the L^2-risk have been obtained in [6]. In particular, it is shown in [6] that

$$R_\varepsilon^2 = \rho_\varepsilon(p, L)(1 + o(1)),$$

where

$$\rho_\varepsilon(p, L) = \frac{1}{2}\left(\frac{\pi^4 p}{3(p+2)}\right)^{2p/(2p+2)} ((2p+2)L)^{2/(2p+2)}\, \varepsilon^{4p/(2p+2)},$$

and an adaptive penalized blockwise Stein-type estimator $\hat{f}$ was constructed (see [6], Eqs. (3.12) and (5.8)) such that

$$\lim_{\varepsilon\to 0} \sup_{p\in[p_1,p_2],\, L\in[L_1,L_2]} \sup_{f\in\mathcal{F}(p,L)} \frac{E||\hat{f}-f||^2}{\rho_\varepsilon(p,L)} = 1$$

for any $0 < p_1 < p_2 < \infty$ and $0 < L_1 < L_2 < \infty$. Note that, in contrast to adaptive minimax separation rates, there is no price to pay for adaptation in the corresponding minimax estimation problem under the L^2-risk (i.e., a global measure). The unavoidable logarithmic factor for adaptivity that appears in the minimax separation rates $r_\varepsilon^{ad}(p)$ stated in Theorem 4.3 also appears in various other adaptivity problems, such as minimax signal detection (see [13]) and minimax estimation under the L^2-risk (see [24]) in some ill-posed inverse problems. It also resembles the minimal price one needs to pay for adaptation in minimax estimation under the l^2-risk (i.e., a local or pointwise measure) that has been observed in [18], [2], [19] and [23] in the case of Lipschitz and Sobolev balls and, more recently, in [3] and [1] in the case of Besov balls.

5. APPENDIX: PROOFS

For simplicity in the calculations, we omit the factors L^{-1} and π^{-1} in (4.9) and (4.10), respectively. In other words, from now on, we work with $a_\nu = (j+1)^p(l+1)^p$ and $\sigma_\nu = (j+l+1)^{1/2}$, $p > 0$, $\nu \in \Gamma$. The final results can be obtained on rescaling by using Remark 4.1.

5.1. Proof of Theorem 4.2

It follows from Theorem 4.1 that the efficiency in the detection problem under consideration is determined by the asymptotics of the quantity

$$u_\varepsilon^2 = \left(\frac{r_\varepsilon}{\varepsilon}\right)^4 \frac{J_0}{2J_1^2},$$

where

$$J_1 = J_1(A) = \sum_{\nu\in\Gamma} \sigma_\nu^4 (1 - Aa_\nu^2)_+, \qquad J_2 = J_2(A) = A\sum_{\nu\in\Gamma} a_\nu^2 \sigma_\nu^4 (1 - Aa_\nu^2)_+,$$
$$J_0 = J_0(A) = \sum_{\nu\in\Gamma} \sigma_\nu^4 (1 - Aa_\nu^2)_+^2 = J_1 - J_2,$$

for $\nu \in \Gamma$ given by (2.3). Moreover, the quantity $A = A_\varepsilon \to 0$ is determined by the relation

$$r_\varepsilon^2 = A\frac{J_1}{J_2}. \tag{5.1}$$

In order to study the asymptotics of u_ε, we are interested in the asymptotics of the functions $J_i(A)$, $i = 0, 1, 2$, as $A \to 0$. We first, however, start with the asymptotics of the following function

$$I(A) = \sum_{\{\nu:\ Aa_\nu^2 \le 1\}} \sigma_\nu^4, \quad \nu \in \Gamma. \tag{5.2}$$

Proposition 5.1. *Let $I(A)$ be defined as in* (5.2)*. Then, as $A \to 0$,*

$$I(A) \sim \frac{2B}{3} A^{-3/(2p)}, \tag{5.3}$$

where $B = \sum_{m=1}^\infty m^{-3} = \zeta(3)$ (cf. Theorem 4.2).

Proof. Set $j+1=m$, $l+1=n$, $H=[m^{-1}A^{-1/(2p)}]$ and $H_1=[A^{-1/(4p)}]$, where $[t]$ is the integer part of t. Consider the set

$$C_{m,n,p,A}=\big\{(m,n)\colon m\geq 1, n\geq 1, (mn)^{2p}\leq A^{-1}\big\}.$$

Then, we have

$$\begin{aligned}I(A)&=\sum_{(m,n)\in C_{m,n,p,A}}(m+n-1)^2\\&=2\sum_{m=1}^{H_1}\sum_{n=1}^{H}(m+n-1)^2-\sum_{m=1}^{H_1}\sum_{n=1}^{H_1}(m+n-1)^2\\&=\frac{1}{3}\sum_{m=1}^{H_1}\big((m-1+H)(m+H)(2m-1+2H)-(m-1)m(2m-1)\big)+O(A^{-1/p})\\&=2\sum_{m=1}^{H_1}(Hm^2+H^2m-Hm-H^2/2+H^3/3+H/6)+O(A^{-1/p}).\end{aligned}$$

We now indicate the asymptotics of the terms in the last sum. Observe that

$$H=\frac{1}{m}A^{-1/(2p)}+\alpha(A,m),\qquad H_1=A^{-1/(4p)}+\beta(A),$$

where $\alpha(A,m)\in[0,1)$ and $\beta(A)\in[0,1)$. Thus, we have

$$\begin{aligned}\sum_{m=1}^{H_1}Hm^2&=\sum_{m=1}^{H_1}(A^{-1/(2p)}m+\alpha(A,m)m^2)=A^{-1/p}/2+O(A^{-3/(4p)}),\\\sum_{m=1}^{H_1}H^2m&=\sum_{m=1}^{H_1}(A^{-1/(2p)}/m+\alpha(A,m))^2\,m\sim A^{-1/p}\log(A^{-1/(4p)})=\frac{A^{-1/p}\log(A^{-1})}{4p},\\\sum_{m=1}^{H_1}Hm&=\sum_{m=1}^{H_1}(A^{-1/(2p)}+\alpha(A,m)m)\sim A^{-3/(4p)},\\\sum_{m=1}^{H_1}H^2/2&=\frac{1}{2}\sum_{m=1}^{H_1}(A^{-1/(2p)}/m+\alpha(A,m))^2\asymp A^{-1/p},\\\sum_{m=1}^{H_1}H^3/3&=\frac{A^{-3/(2p)}}{3}\sum_{m=1}^{H_1}m^{-3}+O(A^{-1/p}),\\\sum_{m=1}^{H_1}H/6&=\sum_{m=1}^{H_1}(A^{-1/(2p)}/m+\alpha(A,m))/6=o(A^{-1/p}).\end{aligned}$$

Therefore

$$I(A)=\frac{2A^{-3/(2p)}}{3}\sum_{m=1}^{H_1}m^{-3}+\frac{A^{-1/p}\log(A^{-1})}{2p}+O(A^{-1/p})\sim\frac{2B}{3}A^{-3/(2p)}.$$

The proposition now follows. □

Let us now return to the asymptotics of $J_i(A)$, $i=0,1,2$, as $A\to 0$. Introduce the following function

$$F(t)=I(t^{-1})=\sum_{(m,n)\in C_{m,n,p,t^{-1}}}(m+n-1)^2,\quad t\geq 0,\tag{5.4}$$

and observe that $F(0) = 0$ and $F(t)$ is nondecreasing in $t \geq 0$. It follows from Proposition 5.1 that

$$F(t) \sim \frac{2B}{3} t^{3/(2p)}, \quad t \to \infty. \tag{5.5}$$

For $T = A^{-1}$, the functions $J_i(A)$, $i = 0, 1, 2$, could be rewritten in the form

$$J_1(A) = \int_0^T (1 - t/T)\, dF(t), \qquad J_2(A) = \int_0^T (t/T - (t/T)^2)\, dF(t),$$

$$J_0(A) = \int_0^T (1 - t/T)^2\, dF(t) = J_1 - J_2.$$

Integrating by parts, we get

$$J_1(A) = T^{-1} \int_0^T F(t)\, dt.$$

In order to study the asymptotics of the above integral, we divide it into the following two parts

$$\int_0^T F(t)\, dt = S_1 + S_2, \qquad S_1 = \int_0^{T^{1/2}} F(t)\, dt, \qquad S_2 = \int_{T^{1/2}}^T F(t)\, dt.$$

Hence it suffices to check that $S_1 = o(S_2)$ and to use the asymptotics (5.5) of the integrand in S_2. It is easily seen that

$$0 \leq S_1 \leq F(T^{1/2}) T^{1/2} \asymp T^{3/(4p)+1/2}$$

and that

$$S_2 \sim \int_{T^{1/2}}^T \left(\frac{2B}{3} t^{3/(2p)}\right) dt \sim \frac{4Bp}{3(2p+3)} T^{1+3/(2p)}.$$

Therefore we get

$$J_1(A) \sim \frac{4Bp}{3(2p+3)} A^{-3/(2p)}. \tag{5.6}$$

Similarly, for $T = A^{-1}$, we get

$$J_2(A) = T^{-1} \int_0^T t\, dF(t) - T^{-2} \int_0^T t^2\, dF(t) = 2T^{-2} \int_0^T tF(t)\, dt$$

$$- T^{-1} \int_0^T F(t)\, dt \sim T^{3/(2p)} \left(\frac{8Bp}{3(4p+3)} - \frac{4Bp}{3(2p+3)}\right) = \frac{4Bp}{(4p+3)(2p+3)} A^{-3/(2p)} \tag{5.7}$$

and that

$$J_0 = J_1 - J_2 \sim \frac{16Bp^2}{3(2p+3)(4p+3)} A^{-3/(2p)}. \tag{5.8}$$

Thus, it follows from (5.1), (5.6), (5.7) and (5.8) that

$$A \sim \frac{3}{4p+3} r_\varepsilon^2, \quad u_\varepsilon^2 \sim r_\varepsilon^{4+3/p} \varepsilon^{-4} \frac{2p+3}{2B} \left(\frac{3}{4p+3}\right)^{1+3/(2p)}. \tag{5.9}$$

Therefore the separation rates r^*_ε are of the form

$$r^*_\varepsilon = \varepsilon^{4p/(4p+3)}. \tag{5.10}$$

In order to get the sharp asymptotics, in view of Theorem 4.1, it is enough to check the condition

$$w_0 = \frac{\max_{\{\nu:\, a_\nu^2 A<1\}} \sigma_\nu^2(1-Aa_\nu^2)}{\sqrt{2\sum_{\{\nu:\, a_\nu^2 A<1\}} \sigma_\nu^4(1-Aa_\nu^2)^2}} = o(1), \quad \nu \in \Gamma.$$

This condition follows directly from the relation

$$\max_{\{\nu:\, a_\nu^2 A<1\}} \sigma_\nu^4 = o(J_0(A)). \tag{5.11}$$

Indeed, for $m = j+1 \in \mathbb{N}$, $n = l+1 \in \mathbb{N}$, the last condition follows from (see (5.8))

$$\sigma_\nu^4 = (m+n-1)^2 < 4A^{-1/p} \ll A^{-3/2p} \asymp J_0(A)$$

as $A \to 0$, for $\nu = (m-1, n-1)$ such that $a_\nu^2 = (mn)^{2p} \le A^{-1}$.

The theorem now follows. □

5.2. Proof of Theorem 4.3

Let p be unknown and consider $p \in \Sigma$, where $\Sigma = [p_{\min}, p_{\max}]$, $0 < p_{\min} < p_{\max} < \infty$, is a compact interval in $(0,\infty)$. Recall the expression $\phi(p)$ from (4.16) and that $Z_+ = \mathbb{N} \cup \{0\}$.

We first obtain the lower bounds. Take a collection p_k, $k \in Z_+$, such that

$$\phi(p_k) = a + k\delta_\varepsilon, \qquad k = 0, 1, \dots, K = K_\varepsilon,$$

with

$$\phi(p_K) = b > a = \phi(p_0) > 0, \qquad \delta = \delta_\varepsilon = \frac{(b-a)}{K} \sim \frac{\log(2)}{\log(\varepsilon^{-1})}.$$

Therefore

$$p_k \in [b^{-1} - 3/4, a^{-1} - 3/4] = \Sigma, \qquad b < 4/3,$$

and

$$\delta_\varepsilon = \phi(p_k) - \phi(p_{k-1}), \qquad p_{k-1} - p_k = (4p_k+3)(4p_{k-1}+3)\delta_\varepsilon/16 \asymp \frac{1}{\log(\varepsilon^{-1})} = o(1).$$

Assume without loss of generality that, uniformly in $p \in \Sigma$,

$$u_\varepsilon(p) \sim \sqrt{d \log\log(\varepsilon^{-1})}, \tag{5.12}$$

where the constant $d > 0$ will be specified below. This corresponds to taking, uniformly in $p \in \Sigma$,

$$r_\varepsilon(p) \sim \big(\varepsilon(d(p)\log\log(\varepsilon^{-1})\,)^{1/4}\big)^{\phi(p)p}, \tag{5.13}$$

where

$$d(p) = da(p), \quad a(p) = \frac{2B}{2p+3}\left(\frac{3}{4}\phi(p)\right)^{-(2p+3)/(2p)}.$$

Take

$$T_k \sim (2r_\varepsilon(p_k))^{-1/p_k}, \quad k = 0, 1, \dots, K_\varepsilon.$$

By construction, we have

$$\begin{aligned} T_k - T_{k-1} &\sim T_{k-1}\Big(2^{p_{k-1}^{-1} - p_k^{-1}}(\varepsilon\sqrt[4]{d\log\log(\varepsilon^{-1})})^{-\delta_\varepsilon} - 1\Big) \\ &= T_{k-1}\big(\exp(\log(2)(1+o(1))) - 1\big) \sim T_{k-1}. \end{aligned}$$

Observe that the function $F(t) = F_p(t)$ defined by (5.4) depends on p. Also, by (5.5), we can write

$$F(t) = F_p(t) = \sum_{\{((i_1+1)(i_2+1))^{2p} \le t,\ i_1, i_2 \in Z_+\}} (i_1 + i_2 + 1)^2$$
$$= \sum_{\{\nu:\ a_\nu^2 \le t\}} \sigma_\nu^4 = F_1(t^{1/p}) \sim \frac{2B}{3} t^{3/(2p)}. \qquad (5.14)$$

Set

$$\Delta_k = \{\nu_0 = (j, l) \in \mathbb{Z}_+ \times \mathbb{Z}_+ \colon T_{k-1} < (j+1)(l+1) \le T_k\}.$$

Take a collection $z_k > 0$, v_{k,ν_0}, $\nu_0 \in \Delta_k$, $k = 1, 2, \dots, K$, such that

$$v_{k,\nu_0} = z_k \begin{cases} \xi_{\nu_0} \sigma_{\nu_0} & \text{if} \quad \nu_0 \in \Delta_k, \\ 0 & \text{otherwise,} \end{cases}$$

with $\xi_{\nu_0} = \pm 1$, $\nu_0 \in \Delta_k$. We then have

$$\sum_{\nu_0 \in \Delta_k} v_{k,\nu_0}^2 \sigma_{\nu_0}^2 = z_k^2 \sum_{\nu_0 \in \Delta_k} \sigma_{\nu_0}^4 = z_k^2 \big(F_1(T_k^2) - F_1(T_{k-1}^2)\big) \sim \frac{7B}{12} z_k^2 T_k^3 = 2 r_\varepsilon^2(p_k) \qquad (5.15)$$

and

$$\sum_{\nu_0 \in \Delta_k} v_{k,\nu_0}^2 \sigma_{\nu_0}^2 a_{\nu_0,p_k}^2 = z_k^2 \sum_{\nu_0 \in \Delta_k} \sigma_{\nu_0}^4 a_{\nu_0,p_k}^2 \le 2 T_k^{2p_k} r_\varepsilon^2(p_k)(1 + o(1)) = \frac{1}{2}(1 + o(1)) < 1. \qquad (5.16)$$

Furthermore, we have

$$u_\varepsilon^2(p_k) \sim d \log\log(\varepsilon^{-1}), \qquad d > 0, \qquad p_k \in \Sigma.$$

Consider now the priors

$$\pi_k = \prod_{\nu_0 \in \Delta_k} (\delta_{z_k \sigma_{\nu_0} e_{\nu_0}} + \delta_{-z_k \sigma_{\nu_0} e_{\nu_0}})/2, \quad k = 1, 2, \dots, K, \qquad \pi = \frac{1}{K} \sum_{k=1}^K \pi_k,$$

where $\{e_{\nu_0}\}_{\nu_0 \in \mathbb{Z}_+ \times \mathbb{Z}_+}$ is the standard basis in the space l^2 that corresponds to sequences indexed by $\nu_0 \in \mathbb{Z}_+ \times \mathbb{Z}_+$, and δ_η is the Dirac mass at the point $\eta \in l^2$. The relations (5.15) and (5.16) imply

$$\pi_k\big(\Theta_\varepsilon(p_k, r_\varepsilon(p_k))\big) = 1, \qquad \pi(\Theta_\varepsilon(\Sigma)) = 1.$$

Let $P_{\pi_k} = E_{\pi_k} P_{\varepsilon,\eta}$, $P_\pi = E_\pi P_{\varepsilon,\eta}$ be the mixtures over the priors. It suffices to check that

$$E_{\varepsilon,0}\big((dP_\pi / dP_{\varepsilon,0} - 1)^2\big) = o(1). \qquad (5.17)$$

Using evaluations similar to Section 5.6 in [13], we have

$$E_{\varepsilon,0}\big((dP_\pi / dP_{\varepsilon,0} - 1)^2\big) = \frac{1}{K^2} \sum_{k=0}^K E_{\varepsilon,0}\big((dP_{\pi_k} / dP_{\varepsilon,0} - 1)^2\big)$$
$$= \frac{1}{K^2} \sum_{k=0}^K \big(E_{\varepsilon,0}(dP_{\pi_k} / dP_{\varepsilon,0})^2 - 1\big)$$
$$\le \frac{1}{K^2} \sum_{k=0}^K \Big(\exp\ \Big(2 \sum_{\nu_0 \in \Delta_k} \sinh^2(z_k^2 \sigma_{\nu_0}^2 / 2\varepsilon^2)\Big) - 1\Big).$$

Note that $\sigma_{\nu_0}^2 = j + l + 1 < T_k$ if $\nu_0 \in \Delta_k$. Therefore, uniformly over $\nu_0 \in \Delta_k$, we have

$$\frac{z_k^2 \sigma_{\nu_0}^2}{\varepsilon^2} < \frac{z_k^2 T_k}{\varepsilon^2} \asymp r_\varepsilon(p_k)^{2+2/p_k} \varepsilon^{-2} \asymp \varepsilon^{2/(4p_k+3)} \big(\log\log(\varepsilon^{-1})\big)^{(2p_k+2)/(4p_k+3)} = o(1),$$

and (since $\sinh^2(z_k^2\sigma_{\nu_0}^2/2\varepsilon^2) \sim z_k^4\sigma_{\nu_0}^4/4\varepsilon^4$)

$$2\sum_{\nu_0\in\Delta_k}\sinh^2(z_k^2\sigma_{\nu_0}^2/2\varepsilon^2) \sim z_k^2\varepsilon^{-4}r_\varepsilon^2(p_k) \sim \frac{24}{7B}2^{3/p_k}r_\varepsilon^{4+3/p_k}(p_k)\varepsilon^{-4}$$

$$\sim \frac{24}{7B}2^{3/p_k}da(p_k)\log\log(\varepsilon^{-1}).$$

One can take $d>0$ such that, for any $p_k\in\Sigma$,

$$d\frac{24}{7B}2^{3/p_k}a(p_k) \le d\frac{16}{7}\max_{k=0,1,\dots,K_\varepsilon}\left\{\frac{4p_k+3}{2p_k+3}\left(\frac{4(4p_k+3)}{3}\right)^{3/(2p_k)}\right\} = d_1 < 1.$$

Then we have

$$E_{\varepsilon,0}\big((dP_\pi/dP_{\varepsilon,0}-1)^2\big) \le \frac{1}{K^2}\sum_{k=0}^{K}\Big(\exp\Big(2\sum_{\nu_0\in\Delta_k}\sinh^2(z_k^2\sigma_{\nu_0}^2/2\varepsilon^2)\Big)-1\Big)$$

$$< \frac{K\log^{d_1}(\varepsilon^{-1})}{K^2} \asymp \frac{\log^{d_1}(\varepsilon^{-1})}{\log(\varepsilon^{-1})} = o(1).$$

We now obtain the upper bounds.

Similarly to the proof of the lower bounds, assume without loss of generality that $u_\varepsilon(p) \sim \sqrt{D\log\log(\varepsilon^{-1})}$, uniformly in $p\in\Sigma$, where the constant $D>0$ will be specified below. This corresponds to (5.13) with d replaced by D, uniformly in $p\in\Sigma$.

In order to evaluate the type I error probability, we consider a different grid with different $K=K_\varepsilon$, i.e.,

$$\phi(p_k) = a + k\delta_\varepsilon, \qquad k=0,1,\dots,K=K_\varepsilon, \qquad \phi(p_K)=b>a>0,$$

where

$$\delta=\delta_\varepsilon=\frac{(b-a)}{K_\varepsilon} \sim \frac{\log(2)}{\log(\varepsilon^{-1})\log\log(\varepsilon^{-1})}.$$

Let us evaluate the exponential moments

$$E_{\varepsilon,0}(\exp(ht_{\varepsilon,c_k})), \quad h>0.$$

Recall that $(y_\nu/\varepsilon) \overset{iid}{\sim} \mathcal{N}(0,1)$, $\nu\in\Gamma$, under P_0. Recall the set $C_{\nu,k}=\{\nu\colon a_{\nu,p_k}\le c_k\}$, $\nu\in\Gamma$, $k=0,1,\dots,K$. Let the family $h=h_\varepsilon$ be taken in such a way that

$$h\max_{\nu\in C_{\nu,k}} w_{\nu,k} = o(1).$$

Then we have

$$E_{\varepsilon,0}(\exp(ht_{\varepsilon,c_k})) = \prod_{\nu\in C_{\nu,k}}\big(\exp(-hw_{\nu,k})E_{\varepsilon,0}\exp(hw_{\nu,k}\xi_\nu^2)\big)$$

$$= \exp\Big(\sum_{\nu\in C_{\nu,k}}(-hw_{\nu,k}-\log(1-2hw_{\nu,k})/2)\Big)$$

$$= \exp\Big(\sum_{\nu\in C_{\nu,k}}h^2w_{\nu,k}^2(1+O(hw_{\nu,k}))\Big) = \exp(h^2/2)(1+o(1)). \tag{5.18}$$

Let $h=H_\varepsilon$. Then, for $k=0,1,\dots,K_\varepsilon$, we have

$$h\max_{\nu\in C_{\nu,k}} w_{\nu,k} = H_\varepsilon\frac{\max_{\nu\in C_{\nu,k}}\sigma_\nu^2}{\big(2\sum_{\nu\in C_{\nu,k}}\sigma_\nu^4\big)^{1/2}} < \frac{H_\varepsilon c_k^{1/p_k}}{\big(4Bc_k^{3/p_k}(1+o(1))/3\big)^{1/2}}$$

$$\asymp H_\varepsilon r_\varepsilon^{1/(2p_k)}(p_k) \asymp \varepsilon^{2/(3+4p_k)}\big(\log\log(\varepsilon^{-1})\big)^{2(1+p_k)/(3+4p_k)} = o(1),$$

and by (5.18), for any $k = 0, 1, \dots, K_\varepsilon$, we have

$$P_{\varepsilon,0}(t_{\varepsilon,c_k} > H_\varepsilon) \le \frac{E_{\varepsilon,0}(\exp(H_\varepsilon t_{\varepsilon,c_k}))}{\exp(H_\varepsilon^2)} \sim \exp(H_\varepsilon^2/2 - H_\varepsilon^2) = \exp(-H_\varepsilon^2/2) = K_\varepsilon^{-2}.$$

This implies that, for the type I error probability,

$$\alpha(\psi_\varepsilon) \le \sum_{k=0}^{K_\varepsilon} P_{\varepsilon,0}(t_{\varepsilon,c_k} > H_\varepsilon) \le K_\varepsilon^{-1}(1 + o(1)) \to 0.$$

Let us evaluate the type II error probability for

$$\eta \in \Theta_\varepsilon(\Sigma) = \bigcup_{p\in\Sigma} \Theta_{\varepsilon,p}(r_\varepsilon(p)).$$

There exists p such that $\eta \in \Theta_{\varepsilon,p}(r_\varepsilon(p))$, $p_k \le p \le p_{k-1}$. Observe that

$$\beta_\varepsilon(\eta, \psi_\varepsilon) \le \min_{0\le k\le K_\varepsilon} P_{\varepsilon,\eta}(t_{\varepsilon,c_k} \le H_\varepsilon).$$

Denote $h_{\varepsilon,c_k} = E_{\varepsilon,\eta}(t_{\varepsilon,c_k})$. We then have

$$h_{\varepsilon,c_k} = \varepsilon^{-2} \sum_{\nu\in C_{\nu,k}} w_{\nu,k}\eta_\nu^2; \qquad \mathrm{Var}_{\varepsilon,\eta}(t_{\varepsilon,c_k}) = 1 + 4\varepsilon^{-2} \sum_{\nu\in C_{\nu,k}} w_{\nu,k}^2\eta_\nu^2 = 1 + O(h_{\varepsilon,c_k}). \tag{5.19}$$

Let us now evaluate

$$h_{\varepsilon,c_k} = \frac{1}{\varepsilon^2\sqrt{2F_{p_k}(c_k^2)}} \sum_{\nu\in C_{\nu,k}} \sigma_\nu^2\eta_\nu^2,$$

where the function

$$F_p(c) = \sum_{\{\nu:\, a_{\nu,p}^2 \le c\}} \sigma_\nu^4$$

is of the form (5.14) with asymptotics given by (5.14) as well. Observe that $a_{\nu,p_k} = a_{\nu,p}^{p_k/p}$ and hence,

$$\sum_{\nu\in C_{\nu,k}} \sigma_\nu^2\eta_\nu^2 = \sum_{\nu\in\Gamma} \sigma_\nu^2\eta_\nu^2 - \sum_{\{\nu:\, a_{\nu,p_k} > c_k\}} \sigma_\nu^2\eta_\nu^2 \ge r_\varepsilon^2(p) - c_k^{-2p/p_k} = r_\varepsilon^2(p)\Big(1 - \frac{1}{r_\varepsilon^2(p)c_k^{2p/p_k}}\Big).$$

Because $a(p)^{p\phi(p)/4}$ satisfy the Lipschitz continuity, we have by (5.13) with d replaced by D

$$\frac{r_\varepsilon(p_k)}{r_\varepsilon(p)} = (1 + O(\delta_\varepsilon)) \exp\Big((p\phi(p) - p_k\phi(p_k))\Big(\log(\varepsilon^{-1}) + \big(\log(D^{-1}) - \log\log\log(\varepsilon^{-1})\big)/4\Big)\Big)$$

and

$$0 \le p\phi(p) - p_k\phi(p_k) = \frac{3}{4}(\phi(p_k) - \phi(p)) \le \frac{3}{4}\delta_\varepsilon \sim \frac{3\log(2)}{4\log(\varepsilon^{-1})\log\log(\varepsilon^{-1})}.$$

We get

$$1 \le \frac{r_\varepsilon(p_k)}{r_\varepsilon(p)} \le 2^{3/(4\log\log(\varepsilon^{-1}))}(1 + o(1)) = 1 + o(1), \tag{5.20}$$

and, moreover, we have

$$\frac{p}{p_k} = 1 + \frac{\Delta p}{p_k}, \qquad 0 \le p - p_k = \Delta p \le p_{k-1} - p_k \asymp \frac{1}{\log(\varepsilon^{-1})\log\log(\varepsilon^{-1})}.$$

Thus we have

$$c_k \le c_k^{p/p_k} = c_k \cdot c_k^{\Delta p/p_k} = c_k(1 + o(1)), \qquad r_\varepsilon(p)c_k^{p/p_k} = 2(1 + o(1)).$$

These relations and (5.20) imply, for $\varepsilon > 0$ small enough,

$$\sum_{\nu \in C_{\nu,k}} \sigma_\nu^2 \eta_\nu^2 \geq r_\varepsilon^2(p_k)\big(1 - 1/4(1+o(1))\big) = r_\varepsilon^2(p_k)(3/4 + o(1)).$$

Therefore, for large enough $D > 0$, we have, for all $k = 1, 2, \dots, K_\varepsilon$,

$$h_{\varepsilon,c_k} \geq \frac{r_\varepsilon^2(p_k)(3/4+o(1))}{\varepsilon^2\sqrt{2F_{p_k}(c_k^2)}} \sim \frac{r_\varepsilon^2(p_k)}{\varepsilon^2 c_k^{3/(2p_k)}}\left(\frac{3\sqrt{3}}{8\sqrt{B}} + o(1)\right)$$

$$\sim \big(Da(p_k)\log\log(\varepsilon^{-1})\big)^{1/2} 2^{-3/(2p_k)}\left(\frac{3\sqrt{3}}{8\sqrt{B}} + o(1)\right) > 2H_\varepsilon \sim 4\sqrt{\log\log(\varepsilon^{-1})}. \qquad (5.21)$$

It follows from (5.19) and (5.21) that, for k such that $\eta \in \Theta_{\varepsilon,p}(r_\varepsilon(p))$, $p_k \leq p \leq p_{k-1}$, one has

$$P_{\varepsilon,\eta}(t_{\varepsilon,c_k} \leq H_\varepsilon) = P_{\varepsilon,\eta}\big(h_{\varepsilon,c_k} - t_{\varepsilon,c_k} \geq h_{\varepsilon,c_k} - H_\varepsilon\big) \leq \frac{\mathrm{Var}_{\varepsilon,\eta}(h_{\varepsilon,c_k} - t_{\varepsilon,c_k})}{(h_{\varepsilon,c_k} - H_\varepsilon)^2}$$

$$< \frac{4\,\mathrm{Var}_{\varepsilon,\eta}(t_{\varepsilon,c_k})}{h_{\varepsilon,c_k}^2} = \frac{4(1+O(h_{\varepsilon,c_k}))}{h_{\varepsilon,c_k}^2} < \frac{1}{4\log\log(\varepsilon^{-1})} + o_\eta(1) = o(1). \qquad (5.22)$$

Hence for any $\eta \in \Theta_\varepsilon(\Sigma)$, one has

$$\beta_\varepsilon(\eta, \psi_\varepsilon) < 1/(4\log\log(\varepsilon^{-1})) + o_\eta(1) = o(1),$$

$$\beta_\varepsilon(\Theta_\varepsilon(\Sigma), \psi_\varepsilon) = \sup_{\eta \in \Theta_\varepsilon(\Sigma)} \beta_\varepsilon(\eta, \psi_\varepsilon) \leq \frac{1}{4\log\log(\varepsilon^{-1})} + \sup_{\eta \in \Theta_\varepsilon(\Sigma)} o_\eta(1),$$

where, by (5.22), (4.14) and (5.19) for any $\eta \in \Theta_\varepsilon(\Sigma)$

$$o_\eta(1) = \frac{4\varepsilon^{-2}\sum_{\nu \in C_{\nu,k}} w_{\nu,k}^2 \eta_\nu^2}{(\varepsilon^{-2}\sum_{\nu \in C_{\nu,k}} w_{\nu,k}\eta_\nu^2)^2} \leq \frac{4\max_{\nu \in C_{\nu,k}} w_{\nu,k}}{h_{\varepsilon,c_k}} < \frac{2}{H_\varepsilon} \sim \big(\log\log(\varepsilon^{-1})\big)^{-1/2}.$$

This completes the proof of part (b) of the theorem.

Part (c) of the theorem follows immediately from parts (a) and (b) of the theorem and (4.11). The theorem now follows. □

ACKNOWLEDGMENTS

We are grateful to the AE and the referee for their useful comments and recommendations.

Research of the first author was partially supported by RFBR grant 11-01-00577 and by the grant NSh–4472.2010.1.

REFERENCES

1. N. Bochkina and T. Sapatinas, "Minimax Rates of Convergence and Optimality of Bayes Factor Wavelet Regression Estimators under Pointwise Risks", Statistica Sinica **19**, 1389–1406 (2009). (Referred (online) Supplement: Statistica Sinica **19**, S21–S37 (2009).)
2. L. D. Brown and M. G. Low, "A Constraint Risk Inequality with Applications to Nonparametric Functional Estimation", Ann. Statist. **24**, 2524–2535 (1996).
3. T. T. Cai, "Rates of Convergence and Adaptation over Besov Spaces under Pointwise Risk", Statistica Sinica **13**, 881–902 (2003).
4. E. J. Candès and D. L. Donoho, "Recovering Edges in Ill-Posed Inverse Problems: Optimality of Curvelet Frames", Ann. Statist. **30**, 784–842 (2002).
5. L. Cavalier, "Nonparametric Statistical Inverse Problems", Inverse Problems **24**, 3, Article ID 034004 (2008).
6. L. Cavalier and A. B. Tsybakov, "Sharp Adaptation for Inverse Problems with Random Noise", Probab. Theory and Rel. Fields **123**, 323–354 (2002).
7. S. R. Deans, *The Radon Transform and Some of its Applications* (Wiley, New York, 1983).

8. S. Efromovich and A. Samarov, "Asymptotic Equivalence of Nonparametric Regression and White Noise Models has Its Limits", Statist. and Probab. Letters **28**, 143–145 (1996).
9. G. T. Herman, *Image Reconstructions from Projections: The Fundamentals of Computerized Tomography* (Academic Press, New York, 1980).
10. Yu. I. Ingster, "Asymptotically Minimax Hypothesis Testing for Nonparametric Alternatives. I, II, III", Math. Methods Statist. **2**, 85–114, 171–189, 249–268 (1993).
11. Yu. I. Ingster and I. A. Suslina, *Nonparametric Goodness-of-Fit Testing under Gaussian Model*, in *Lectures Notes in Statistics* (Springer, New York, 2003), Vol. 169.
12. Yu. I. Ingster and I. A. Suslina, "Nonparametric Hypothesis Testing for Small Type I Errors. I", Math. Methods Statist. **13**, 409–459 (2004).
13. Yu. I. Ingster, T. Sapatinas, and I. A. Suslina, *Minimax Signal Detection in Ill-Posed Inverse Problems*, Preprint (2011), arXiv:1001.1853v3 [math.ST].
14. I. M. Johnstone and B. W. Silverman, "Speed of Estimation in Positron Emission Tomography and Related Inverse Problems", Ann. Statist. **18**, 251–280 (1990).
15. G. Kerkyacharian, G. Kyriazis, E. Le Pennec, P. Petrushev, and D. Picard, "Inversion of noisy Radon transform by SVD based needlets", Appl. and Comput. Harmonic Analysis **28**, 24–45 (2010).
16. G. Kerkyacharian, E. Le Pennec, and D. Picard, *Radon Needlet Thresholding*, Preprint (2009), arXiv:0908.2514v1 [math.ST].
17. A. Korostelev and A. Tsybakov, *Minimax Theory of Image Reconstruction* (Springer, New York, 1993).
18. O. V. Lepski, "On a Problem of Adaptive Estimation in White Gaussian Noise", Theory Probab. Appl. **35**, 454–466 (1990).
19. O. V. Lepski and V. G. Spokoiny, "Optimal Pointwise Adaptive Methods in Nonparametric Estimation", Ann. Statist. **25**, 2512–2546 (1997).
20. F. Natterer, *The Mathematics of Computerized Tomography* (SIAM, Philadelphia, 2001).
21. M. Nussbaum, "Spline Smoothing in Regression Models and Asymptotic Efficiency in L_2", Ann. Statist. **13**, 984–997 (1985).
22. V. G. Spokoiny, "Adaptive Hypothesis Testing Using Wavelets", Ann. Statist. **24**, 2477–2498 (1996).
23. A. B. Tsybakov, "Pointwise and Sup-Norm Sharp Adaptive Estimation of Functions on the Sobolev Classes", Ann. Statist. **26**, 2420–2469 (1998).
24. A. B. Tsybakov, "On the Best Rate of Adaptive Estimation in Some Inverse Problems", C. R. Acad. Sci. Paris, Sér. I **330**, 835–840 (2000).