ТЕОРИЯ ВЕРОЯТНОСТЕЙ
Том 48 **И ЕЕ ПРИМЕНЕНИЯ** **Выпуск 3**
2003

 CAVALIER L. [*], ***GOLUBEV Y.*** [*],
LEPSKI O. [*], ***TSYBAKOV A.*** [**]

BLOCK THRESHOLDING AND SHARP ADAPTIVE ESTIMATION IN SEVERELY ILL-POSED INVERSE PROBLEMS[1)]

Рассматривается проблема решения линейных операторных уравнений, полученных из наблюдений с шумом, в предположении, что сингулярные значения оператора экспоненциально убывают и что преобразование Фурье соответствующего решения также экспоненциально гладко. Мы предлагаем оценку решения, основанную на скользящем варианте блокового порога в пространстве коэффициентов Фурье. Показано, что эта оценка может быть быстро адаптировано к неизвестной степени гладкости решения.

Ключевые слова и фразы: линейное операторное уравнение, некорректно поставленные задачи, наблюдения с шумом.

1. Introduction. The problem of solving linear operator equations from noisy observations has been extensively studied in the literature. Among the first to develop a statistical approach to this problem were Sudakov and Khalfin [17] and Bakushinskii [1]. For a survey of recent results we refer to Mathé and Pereverzev [14], Goldenshluger and Pereverzev [7], Cavalier and Tsybakov [3].

A usual statistical framework in this context is as follows. Let $K\colon H\to H$ be a known linear operator on a Hilbert space H with inner product $(\cdot,\cdot)$ and norm $\|\cdot\|$. The problem is to estimate an unknown function $f\in H$ from indirect observations

$$Y(g)=(Kf,g)+\varepsilon\xi(g),\qquad g\in H, \tag{1}$$

where $0<\varepsilon<1$ and $\xi(g)$ is a zero-mean Gaussian random process indexed by H on a probability space $(\Omega,\mathscr{A},\mathbf{P})$, such that $\mathbf{E}\{\xi(g)\xi(v)\}=(g,v)$ for any $g,v\in H$, where $\mathbf{E}$ is the expectation w.r.t. $\mathbf{P}$. Relation (1) defines a Gaussian white noise model.

[*] CMI, Université Aix-Marseille 1, 39 rue F. Joliot-Curie, F-13453 Marseille Cedex, France.
[**] Laboratoire de Probabilités et Modèles Aléatoires, Université Paris 6, 4 pl. Jussieu, BP 188, F-75252 Paris Cedex 05, France.

Instead of dealing with all the observations $\{Y(g),\ g \in H\}$, it is usually sufficient to consider a sequence of values $\{Y(g_k)\}_{k=1}^{\infty}$, for some orthonormal basis $\{g_k\}_{k=1}^{\infty}$. The corresponding random errors $\xi(g_k) = \xi_k$ are i.i.d. standard Gaussian random variables.

We assume that the basis $\{g_k\}$ is such that $(Kf, g_k) = b_k\theta_k$, where $b_k \neq 0$ are real numbers and $\theta_k = (f, \varphi_k)$ are the Fourier coefficients of f w.r.t. some orthonormal basis $\{\varphi_k\}$ (not necessarily $\varphi_k = g_k$). A typical example when it occurs is that the operator K admits a *singular value decomposition*:

$$K\varphi_k = b_k g_k, \qquad K^* g_k = b_k \varphi_k, \tag{2}$$

where K^* is the adjoint of K, b_k are singular values, $\{g_k\}$ is an orthonormal basis in Range(K) and $\{\varphi_k\}$ is the corresponding orthonormal basis in H.

Under these assumptions, one gets a discrete sequence of observations derived from (1):

$$y_k = b_k\theta_k + \varepsilon\xi_k, \qquad k = 1, 2, \dots, \tag{3}$$

where $y_k = Y(g_k)$ and ξ_i are i.i.d. standard Gaussian random variables. The problem of estimating f reduces to estimation the sequence $\{\theta_k\}_{k=1}^{\infty}$ from observations (3). The model (3) also describes other problems such as the estimation of a signal from direct observations with correlated data (see Johnstone [11]).

Let $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \dots)$ be an estimator of $\theta = (\theta_1, \theta_2, \dots)$ based on the data (3). Then f is estimated by $\hat{f}c = \sum_k \hat{\theta}_k \varphi_k$. The mean integrated squared error of the estimator $\hat{f}$ is

$$\mathbf{E}\|\hat{f} - f\|^2 = \mathbf{E}_\theta \sum_{k=1}^{\infty} (\hat{\theta}_k - \theta_k)^2 \stackrel{\text{def}}{=} R^\varepsilon(\hat{\theta}, \theta), \tag{4}$$

where $\mathbf{E}_\theta$ denotes the expectation w.r.t. the distribution of the data in the model (3).

In this paper we consider the problem of estimation of θ in the model (3) using the mean-squared risk (4).

One can characterize linear inverse problems by the difficulty of the operator, i.e., with our notations, by the behavior of b_k's. If $b_k \to 0$, as $k \to \infty$, the problem is ill-posed. An inverse problem will be called *softly ill-posed* if the sequence b_k tends to 0 at a polynomial rate in k and it will be called *severely ill-posed* if

$$\lim_{k\to\infty} \frac{1}{k} \log \frac{1}{b_k} = c,$$

for some $0 < c < \infty$. Thus, the problem is severely ill-posed if, in the main term, b_k tends to 0 exponentially in k.

An important element of the model is the prior information about θ. Successful estimation of a sequence θ is possible only if its elements θ_k tend to zero sufficiently fast, as $k \to \infty$, which means that f is sufficiently smooth. A standard assumption on the smoothness of f is to suppose that θ belongs to an ellipsoid

$$\Theta = \left\{ \theta\colon \sum_{k=1}^{\infty} a_k^2 \theta_k^2 \leqslant L \right\},$$

where $a = \{a_k\}$ is a positive sequence that tends to infinity, and $L > 0$. Special cases of Θ are the Sobolev balls and the classes of analytic functions, corresponding to a_k's increasing as a polynomial in k and as an exponential in k, respectively.

Thus appears a natural classification of different cases in the study of linear inverse problems. Regarding the difficulty of the operator described in terms of b_k's and the smoothness assumptions described in terms of a_k's, one obtains the following three typical cases.

1°. *Softly ill-posed problems*: b_k's are polynomial and a_k's are general (usually polynomial or exponential). These problems have been studied by many authors, and they are essentially similar to estimation of derivatives of smooth functions. Sharp adaptive estimators for a general framework are given by Cavalier and Tsybakov [3], and by Cavalier, Golubev, Picard, and Tsybakov [2].

2°. *Severely ill-posed problems with log-rates*: b_k's are exponential and a_k's are polynomial. This case is highly degenerate in the sense that the variance of the optimal estimators is asymptotically negligible as compared to their bias. The optimal rates of convergence are very slow (logarithmic) and sharp adaptation can be attained on a simple projection estimator ([4], [8]).

3°. $2\exp$-*severely ill-posed problems*: b_k's are exponential and a_k's are exponential too (the abbreviation «$2\exp$» stands for «two exponentials»). These problems will be studied here. They are characterized by some unusual phenomena. Golubev and Khasminskii [9] proved that $2\exp$ problems admit fast optimal rates converging to 0 as a power law, despite the «severe» form of the operator. They also showed that sharp minimax estimators for these problems are nonlinear, unlike all other known cases where sharp minimaxity is explored. Also the adaptation issue turns out to be nonstandard here. As shown by Tsybakov [19], there is a logarithmic deterioration in the best rate of convergence for adaptive estimation under the L_2-risk. In other words, here one has to pay a price for L_2-adaptation, while this is not the case for the inverse problems described in 1° and 2°: there the L_2-adaptation is possible without any loss, and even the exact constants are preserved.

Since the ellipsoid Θ with exponential a_k's corresponds to analytic functions, the $2\exp$ framework can be viewed as an analogue of convolution through a super-smooth filter (described by exponential b_k's), with an analytical function f to reconstruct.

There is an important reason why the 2 exp setup is of interest. In the study of inverse problems, a standard assumption is to connect the smoothness of the underlying function to the smoothness of the operator. Roughly, if a function is observed through a very smooth filter, then the function itself has to be very smooth. A formalization of this idea can be found, for example, in the well-known Hilbert scale approach to inverse problems (see [7], [13]–[15]).

Nonadaptive minimax estimation for some inverse problems different from (3) but characterized by a similar «two exponentials» behavior has been analyzed by Ermakov [6], Pensky and Vidakovic [16]), Efromovich and Koltchinskii [5].

In this paper we study estimation for the 2 exp framework when the ellipsoid Θ is not known, but we know only that a_k's are exponential. We propose an adaptive estimator which attains optimal rates (up to an inevitable logarithmic factor deterioration) simultaneously on all the ellipsoids with exponential a_k's. Moreover, we show that the estimator is sharp adaptive, i.e., it cannot be improved to within a constant. This generalizes the result of Tsybakov [19] about the optimal rate of adaptation for 2 exp problems. The construction of our adaptive estimator is based on a block thresholding (cf. [10] or [3] for the inverse problems setting). A difference from those papers is that, in order to get sharp optimality in our case, we need a «running» block estimator rather than an estimator with fixed blocks.

Let us give some examples of severely ill-posed inverse problems related to partial differential equations.

E x a m p l e 1. Consider the Dirichlet problem for the Laplace equation on a circle of radius 1:

$$\Delta u = 0, \quad u(1,\varphi) = f(\varphi), \qquad \varphi \in [0,2\pi], \quad 0 \leqslant r \leqslant 1, \tag{5}$$

where Δ is the Laplace operator, $u(r,\varphi)$ is a function in polar coordinates $r \geqslant 0, \varphi \in [0,2\pi]$, and f is a 2π-periodic function in $L_2[0,2\pi]$. It is well-known that the solution of (5) is

$$u_f(r,\varphi) = \frac{\theta_0}{\sqrt{2\pi}} + \frac{1}{\sqrt{\pi}} \sum_{k=1}^{\infty} r^k [\theta_k \cos(k\varphi) + \theta_{-k} \sin(k\varphi)], \tag{6}$$

where θ_k are the Fourier coefficients of f. Assume that f is not known, but one can observe the solution $u_f(r,\varphi)$ on the circle of radius $r_0 < 1$ in a white Gaussian noise:

$$dY(\varphi) = u_f(r_0,\varphi)d\varphi + \varepsilon dW(\varphi), \qquad \varphi \in [0,2\pi], \tag{7}$$

where W is a standard Wiener process on $[0, 2\pi]$ and $0 < \varepsilon < 1$. The problem is to estimate the boundary condition f based on the observation of a trajectory $\{Y(\varphi), \varphi \in [0, 2\pi]\}$. Substituting (6) in (7), multiplying (7) by the trigonometric basis functions and integrating over $[0, 2\pi]$ we get the infinite sequence of observations $y_k = b_k\theta_k + \varepsilon\xi_k$, $k \in \mathbf{Z}$, where $b_k = r_0^{-|k|}$ and ξ_k are i.i.d. $\mathscr{N}(0, 1)$ random variables. By renumbering the indices from $k \in \mathbf{Z}$ to $k \in \mathbf{N}$ we get a particular case of the model (3). This problem is severely ill-posed since $b_k \to 0$ exponentially fast as $k \to \infty$.

Example 2. Consider the following Cauchy problem for the Laplace equation:

$$\Delta u = 0, \quad u(x, 0) = 0, \quad \frac{\partial}{\partial y} u(x, y)\Big|_{y=0} = g(x), \tag{8}$$

where $u(x, y)$ is defined for $x \in \mathbf{R}$, $y \geqslant 0$, and the initial condition g is a 1-periodic function on $\mathbf{R}$.

Suppose that we do not know g but we have in our disposal the noisy observations $\{Y(x), x \in [0, 1]\}$, where Y is the random process defined by

$$dY(x) = g(x)\,dx + \varepsilon\,dW(x), \qquad x \in [0, 1]. \tag{9}$$

Here W is the standard Wiener process on $[0, 1]$. The problem is to estimate the solution $f(x) \stackrel{\text{def}}{=} u_g(x, y_0)$ of (8) at a given $y_0 > 0$, based on these observations. Since g is 1-periodic, f is also 1-periodic. Denoting θ_k the Fourier coefficients of f, one can find that, given (9), the following sequence of observations is available: $y_k = b_k\theta_k + \varepsilon\xi_k$, where ξ_k are i.i.d. $\mathscr{N}(0, 1)$ random variables and $b_k \sim k\exp(-\beta y_0 k)$ as $k \to \infty$, for some $\beta > 0$ (see [8] for more details).

2. Setting of the problem. From now on we assume that the observations have the form (3), where ξ_k are i.i.d. $\mathscr{N}(0, 1)$ random variables and the values b_k are defined by

$$b_k^{-2} = r_k \exp(\rho k) \tag{10}$$

with $\rho > 0$ and a positive sequence r_k varying slower than an exponential as $k \to \infty$. Such a definition of b_k covers the examples considered above, whereas considering the squared values of b_k's reflects the fact that the results will be insensitive to the signs.

We assume that r_k is subexponential in the sense of the following definition.

Definition 1. A sequence $\{r_k\}_{k=1}^{\infty}$ is called subexponential if $r_k > 0$ for all k and there exist constants $C_* < \infty$ and $\mu \in (0, 1]$ such that

$$\left|\frac{r_{k+1}}{r_k} - 1\right| \leqslant \frac{C_*}{k^{\mu}}, \qquad k = 1, 2, \ldots. \tag{11}$$

The class of subexponential sequences is rather large, including polynomial, logarithmic and other sequences. It is easy to see that a subexponential sequence r_k satisfies

$$\begin{aligned} a\exp(-ck^{1-\mu}) &\leqslant r_k \leqslant a'\exp(c'k^{1-\mu}) && \text{if} \quad 0<\mu<1, \\ ak^{-c} &\leqslant r_k \leqslant a'k^{c'} && \text{if} \quad \mu=1, \end{aligned} \tag{12}$$

with some positive finite constants a, a', c, and c'.

We will assume that θ belongs to an ellipsoid

$$\Theta(\alpha, L) = \left\{\theta\colon \sum_{k=1}^{\infty} q_k \exp(\alpha k)\, \theta_k^2 \leqslant L\right\}, \tag{13}$$

where q_k is a subexponential sequence and $\alpha > 0$, $L > 0$ are finite constants. In order to shed some light on the estimation of θ in this setup, consider a simple projection estimator $\widetilde{\theta} = (\widetilde{\theta}_1, \widetilde{\theta}_2, \ldots)$ with bandwidth $W \in \mathbf{N}$, i.e.,

$$\widetilde{\theta}_k = \begin{cases} b_k^{-1} y_k, & k \leqslant W, \\ 0, & k > W. \end{cases}$$

The maximal risk of this estimator over $\Theta(\alpha, L)$ is bounded from above as follows

$$\begin{aligned} \sup_{\theta\in\Theta(\alpha,L)} R^{\varepsilon}(\widetilde{\theta}, \theta) &= \sup_{\theta\in\Theta(\alpha,L)} \sum_{k>W} \theta_k^2 + \varepsilon^2 \sum_{k=1}^{W} b_k^{-2} \\ &\leqslant L \sum_{k>W} \exp(-\alpha k)\, q_k^{-1} + \varepsilon^2 \sum_{k=1}^{W} \exp(\rho k)\, r_k. \end{aligned} \tag{14}$$

The minimum of the right-hand side of (14) with respect to W is attained for some W depending on ε such that $W \to \infty$ as $\varepsilon \to 0$ (in fact, otherwise the right-hand side of (14) does not tend to 0 as $\varepsilon \to 0$). Using Lemma 2 (see below) to compute the last two sums in (14), we find that as $W \to \infty$, the right hand-side of (14) is approximated by

$$J(W) = \frac{L\exp(-\alpha W)\, q_W^{-1}}{1 - \mathrm{e}^{-\alpha}} + \varepsilon^2\, \frac{\exp(\rho W)\, r_W}{1 - \mathrm{e}^{-\rho}}\ .$$

The minimizer of $J(W)$ gives an approximately optimal bandwidth. Since for subexponential sequences r_k, q_k and large enough W we have $r_{W-1} \approx r_W \approx r_{W+1}$, $q_{W-1} \approx q_W \approx q_{W+1}$, the necessary conditions of a local minimum at W, namely $J(W) < J(W+1)$ and $J(W) < J(W-1)$, can be written in the form

$$L\exp(-\gamma W) < \varepsilon^2\, \mathrm{e}^{\rho} r_W q_W, \qquad L\exp(-\gamma(W-1)) > \varepsilon^2\, \mathrm{e}^{\rho} r_{W-1} q_{W-1},$$

where $\gamma = \rho + \alpha$. It can be shown that all the local minimizers of $J(W)$ provide essentially the same value of J, so that one can take, for example, the smallest local minimizer

$$W(\alpha, L) = \min\Big\{k \in \mathbf{N}\colon L\exp(-\gamma k) < \varepsilon^2 \mathrm{e}^{\rho} r_k q_k\Big\}. \tag{15}$$

In other words, the minimum of the right-hand side of (14) is approximately attained at $W(\alpha, L)$. This yields the following upper bound for the minimax risk:

$$\inf_{\hat\theta} \sup_{\theta\in\Theta(\alpha,L)} R^{\varepsilon}(\hat\theta, \theta) \leqslant C r_{\varepsilon}(\alpha, L),$$

with a constant $C < \infty$ and the rate of convergence

$$r_{\varepsilon}(\alpha, L) = \varepsilon^2 r_{W(\alpha,L)} \exp\left[\rho W(\alpha, L)\right]. \tag{16}$$

Using the argument of [9] it is not difficult to show that the rate of convergence $r_\varepsilon(\alpha, L)$ cannot be improved in a minimax sense. Unfortunately, the minimax approach has a serious drawback: the optimal (or nearly optimal) choice of the bandwidth depends strongly on the parameters of the functional class $\Theta(\alpha, L)$. In the next section we correct this drawback: we propose an adaptive estimator of θ independent of α and L and attaining the rate which is only logarithmically worse than $r_\varepsilon(\alpha, L)$ on any class $\Theta(\alpha, L)$.

3. Adaptive estimator and its optimality. Define the following estimator $\theta^* = (\theta_1^*, \theta_2^*, \ldots)$ based on block thresholding with running blocks

$$\theta_k^* = b_k^{-1} y_k \,\mathbf{I}\left\{|y\|_k^2 \geqslant 2\varepsilon^2 \rho' k\right\}, \qquad k = 1, 2, \ldots, \tag{17}$$

where $\rho' > \rho$ and

$$\|y\|_k^2 = \sum_{s\in\mathbf{N}\colon |k-s|\leqslant N} y_s^2 \tag{18}$$

with an integer $N \geqslant 1$. Here and later $\mathbf{I}\{\cdot\}$ denotes the indicator function.

It will be clear from the proofs that $\theta_k^* = 0$ with a probability close to 1, whenever k does not belong to a «small» neighborhood of the integer W^* defined by

$$W^* = W^*(\alpha, L) = \min\Big\{k \in \mathbf{N}\colon L\exp(-\gamma k) < 2\varepsilon^2 \rho k r_k q_k\Big\}. \tag{19}$$

We will call W^* *the adaptive bandwidth.* Note that $W^*(\alpha, L)$ is smaller than the optimal bandwidth $W(\alpha, L)$ given in (15) for all ε small enough. For instance if $r_k = q_k \equiv 1$ we have as $\varepsilon \to 0$,

$$W(\alpha, L) = \frac{1}{\gamma}\log\frac{L}{\varepsilon^2} + O(1), \tag{20}$$

whereas

$$W^*(\alpha, L) = \frac{1}{\gamma}\,\log\frac{L}{\varepsilon^2} - \frac{1}{\gamma}\,\log\log\frac{L}{\varepsilon^2} + O(1). \tag{21}$$

For general r_k, q_k the closed form expression for W^* is not available, but using (12) one can see that $W^* \asymp \log(1/\varepsilon)$ as $\varepsilon \to 0$. Nevertheless, possible terms of the order $o(\log(1/\varepsilon))$ in the expression for W^* are not negligible since they can affect the rate of convergence (cf. (21)). Note also that the value W^* need not be known for the construction of our estimator.

In this section we establish the exact asymptotics of the minimax adaptive risk. It turns out that this asymptotics is expressed in terms of the value A^* of the following maximization problem:

$$A^* = A^*(\alpha, L) = \max_{\theta\in\Theta_\infty(\alpha,L)} \sum_{k=-\infty}^{\infty} \exp(\rho k)\,\theta_k^2, \tag{22}$$

where

$$\Theta_\infty(\alpha, L) = \left\{\theta \in \ell^2(\mathbf{Z})\colon \sum_{k=-\infty}^{\infty} \theta_k^2 \leqslant 1,\ \sum_{k=-\infty}^{\infty} \exp(\gamma k)\,\theta_k^2 \leqslant E^*(\alpha, L)\right\}, \tag{23}$$

and

$$E^*(\alpha, L) = \frac{L\exp[-\gamma W^*(\alpha, L)]}{2\varepsilon^2\rho\, W^*(\alpha, L)\, r_{W^*(\alpha,L)}\, q_{W^*(\alpha,L)}}. \tag{24}$$

Note that (22)–(23) is a problem of linear programming w.r.t. θ_k^2's and it has a solution belonging to the boundary of $\Theta_\infty(\alpha, L)$. The values $A^*(\alpha, L)$ and $E^*(\alpha, L)$ depend on ε, but the dependence is not strong: they oscillate between two fixed constants as ε varies. In fact, the definition of W^* implies that for any L and α there exist finite positive constants e_1^*, e_2^* such that $e_1^* \leqslant E^*(\alpha, L) \leqslant e_2^*$ for all ε. This implies the existence of finite positive constants a_1^*, a_2^* (depending on L and α) such that $a_1^* \leqslant A^*(\alpha, L) \leqslant a_2^*$ for all ε. In particular, since $\gamma > \rho$, one can take $a_2^* = e_2^*$.

Define

$$\psi_\varepsilon(\alpha, L) = 2A^*(\alpha, L)\,\varepsilon^2\rho\, W^*(\alpha, L)\exp[\rho W^*(\alpha, L)]\, r_{W^*(\alpha,L)}. \tag{25}$$

The next theorem gives a bound for the maximal risk of the estimator θ^* over $\Theta(\alpha, L)$.

Theorem 1. *Assume that* b_k *satisfies* (10) *and that* r_k *and* q_k *are subexponential. Let* θ^* *be the estimator defined by* (17)–(18) *with* $\rho' > \rho$ *and* $N \in \mathbf{N}$. *Then for any* $\alpha > 0$ *and* $L > 0$ *we have*

$$\limsup_{\varepsilon\to 0}\ \sup_{\theta\in\Theta(\alpha,L)} \frac{R^\varepsilon(\theta^*, \theta)}{\psi_\varepsilon(\alpha, L)} \leqslant \frac{\rho'}{\rho} + C\exp\left[-\frac{N}{2}\,\min(\alpha, \rho)\right], \tag{26}$$

where the constant $C < \infty$ *does not depend on* N *and* ρ'.

Note that if the size $2N+1$ of the block is large and the parameter ρ' is close to ρ, then the right-hand side of (26) approaches 1. Alternatively, one can take $N = N_\varepsilon \to \infty$ and $\rho' = \rho'_\varepsilon \to \rho$ as $\varepsilon \to 0$, satisfying appropriate restrictions, which leads to the next result. For $x \geqslant 0$ write $\lceil x \rceil = \min\{n \in \mathbf{N}\colon n > x\}$.

Theorem 2. *Assume that b_k satisfies* (10) *and that r_k and q_k are subexponential. Let θ^* be the estimator defined by* (17) $-$ $-$(18) *with $N = \lceil \sqrt{\log[\log(1/\varepsilon) \vee 1]} \rceil$ and $\rho' = \rho + N^{-1}$. Then for any $\alpha > 0$ and $L > 0$ we have*

$$\limsup_{\varepsilon\to 0} \sup_{\theta\in\Theta(\alpha,L)} \frac{R^\varepsilon(\theta^*, \theta)}{\psi_\varepsilon(\alpha, L)} \leqslant 1. \tag{27}$$

R e m a r k 1. The estimator θ^* is defined as an infinite sequence. It can be proved that, under our assumptions, the number of nonzero components of this sequence is finite almost surely. However, to construct the estimator (17) one has to check the inequality $\|y\|_k^2 \geqslant 2\varepsilon^2\rho' k$ for all k, which is not realizable in practice. It is easy to propose a realizable version: put $\theta_k^* = 0$ for $k > N_{\max}$ with some $N_{\max} = N_{\max}(\varepsilon) \to \infty$ as $\varepsilon \to 0$. Inspection of the proofs shows that to keep Theorems 1 and 2 valid it suffices to take rather small $N_{\max}$, for example, $N_{\max} \sim \log^2(1/\varepsilon)$. Note also that the choice of N, ρ suggested in Theorem 2 is not the only possible one: there exists a variety of similar values (N, ρ) that allow to attain the result of the theorem. These values are described by some technical conditions that we do not include in the theorem but that can be easily extracted from the proof.

We now show that the upper bound (27) is sharp optimal for adaptation. Note first that for every fixed $\alpha_0 > 0$, $L_0 > 0$ there exists an estimator $\hat\theta$ such that

$$\limsup_{\varepsilon\to 0} \sup_{\theta\in\Theta(\alpha_0,L_0)} \frac{R^\varepsilon(\hat\theta, \theta)}{\psi_\varepsilon(\alpha_0, L_0)} = 0.$$

For example, one can take $\hat\theta$ as a projection estimator with the optimal bandwidth corresponding to (α_0, L_0). Thus, $\hat\theta$ gains over θ^* «at a point» (α_0, L_0). On the other hand, there are points (α, L), where θ^* gains over $\hat\theta$. The next theorem shows that if an estimator $\hat\theta$ gains over θ^* at one point (α_0, L_0), there exists another point (α, L), where $\hat\theta$ looses much more than it gains at (α_0, L_0).

Theorem 3. *Assume that b_k satisfies* (10) *and $r_k = q_k = 1$ for all k. Let an estimator $\hat\theta$ be such that, for some $\alpha_0 > 0$, $L_0 > 0$,*

$$\limsup_{\varepsilon\to 0} \sup_{\theta\in\Theta(\alpha_0,L_0)} \frac{R^\varepsilon(\hat\theta, \theta)}{\psi_\varepsilon(\alpha_0, L_0)} < 1. \tag{28}$$

Then there exists $\alpha' > \alpha_0$ such that for all $\alpha > \alpha'$ and all $L > 0$

$$\liminf_{\varepsilon\to 0} \sup_{\theta\in\Theta(\alpha_0,L_0)} \frac{R^\varepsilon(\hat\theta, \theta)}{\psi_\varepsilon(\alpha_0, L_0)} \sup_{\theta\in\Theta(\alpha,L)} \frac{R^\varepsilon(\hat\theta, \theta)}{\psi_\varepsilon(\alpha, L)} = \infty. \tag{29}$$

Theorems 2 and 3 imply in particular that $\psi_\varepsilon(\alpha, L)$ is the adaptive rate of convergence for our problem. This follows from Definition 3 in [18]), with $\sup_{\theta\in\Theta(\alpha,L)} R^\varepsilon(\hat\theta,\theta)$ being viewed as another rate of convergence $S_\varepsilon(\alpha, L)$. In fact, Theorems 2 and 3 give even more than just the rate: they show that an attempt to improve $\psi_\varepsilon(\cdot,\cdot)$ at any point (α_0, L_0) not only in the rate, but also in the constant, leads to a catastrophic behavior in another point. This property is interpreted as sharp adaptive optimality of the rate $\psi_\varepsilon(\cdot,\cdot)$.

Another consequence of these theorems can be called sharp adaptive optimality of the estimator θ^*. One can modify Theorem 3 by expressing the result in terms of the ratio of maximal risks. For any two estimators $\hat\theta_1$ and $\hat\theta_2$ define

$$G_{\alpha,L}\left(\frac{\hat\theta_1}{\hat\theta_2}\right) = \frac{\sup_{\theta\in\Theta(\alpha,L)} R^\varepsilon(\hat\theta_2,\theta)}{\sup_{\theta\in\Theta(\alpha,L)} R^\varepsilon(\hat\theta_1,\theta)}.$$

This value is interpreted as *the gain of* $\hat\theta_1$ *over* $\hat\theta_2$ *at* (α, L). The larger is $G_{\alpha,L}(\hat\theta_1/\hat\theta_2)$, the better is $\hat\theta_1$ as compared to $\hat\theta_2$. It is easy to see that Theorem 3 and (27) imply the following corollary.

Corollary 1. *Under assumptions of Theorem* 3, *let an estimator* $\hat\theta$ *be such that, for some* $\alpha_0 > 0$, $L_0 > 0$,

$$\liminf_{\varepsilon\to 0} G_{\alpha_0,L_0}\left(\frac{\hat\theta}{\theta^*}\right) > 1, \tag{30}$$

where θ^* *is the estimator defined by* (17)–(18) *and satisfying the assumptions of Theorem* 2. *Then there exists* $\alpha > \alpha_0$ *such that for all* $L > 0$ *and all* ε *small enough*

$$G_{\alpha,L}\left(\frac{\theta^*}{\hat\theta}\right) > \ell_\varepsilon G_{\alpha_0,L_0}\left(\frac{\hat\theta}{\theta^*}\right)$$

with $\ell_\varepsilon \to \infty$ *as* $\varepsilon \to 0$.

R e m a r k 2. It follows from (15) and (16) that for $r_k = q_k \equiv 1$ the nonadaptive rate of convergence $r_\varepsilon(\alpha, L)$ is of the order $\varepsilon^{2\alpha/\gamma}$, while (21) and (25) imply that the adaptive rate satisfies $\psi_\varepsilon(\alpha, L) \asymp \varepsilon^{2\alpha/\gamma}\log(1/\varepsilon)^{\alpha/\gamma}$. Thus, one has to pay an extra log-factor for adaptation. This effect is similar to the one established by Lepski [12] in the case of adaptation at a fixed point, and it is due to a nondegenerate asymptotic behavior of the normalized loss of the estimators as $\varepsilon \to 0$. Our problem provides the first example, where such an effect occurs for the L_2-loss and not for the loss at a fixed point.

4. Proofs. In this section we will denote by C finite positive constants that may be different in different occasions.

4.1. Proof of Theorems 1 and 2.

Lemma 1. *Let w_k be a subexponential sequence. Then for any integers T, t such that $t \geqslant T$ and any integer $M < \min(T, \log T)$ we have*

$$\sup_{k\in\mathbf{Z}:\,|k|\leqslant M}\left|\frac{w_{t+k}}{w_t}-1\right|\leqslant \eta_T, \tag{31}$$

where η_T depends only on T and $\eta_T \to 0$ as $T \to \infty$.

P r o o f is straightforward.

Lemma 2. *Let w_k be a subexponential sequence. Then for any $\tau > 0$ as $T \to \infty$*

$$\sum_{k=1}^{T}\exp(\tau k)\,w_k \;=\; (1+o(1))\,\frac{\exp(\tau T)\,w_T}{1-\mathrm{e}^{-\tau}}\,, \tag{32}$$

$$\sum_{k=T}^{\infty}\exp(-\tau k)\,w_k \;=\; (1+o(1))\,\frac{\exp(-\tau T)\,w_T}{1-\mathrm{e}^{-\tau}}\,. \tag{33}$$

P r o o f. Write

$$\sum_{k=1}^{T}\exp(\tau k)\,w_k = \exp(\tau T)\,w_T\sum_{k=0}^{T-1}\exp(-\tau k)\,\frac{w_{T-k}}{w_T}.$$

Set $M = \lceil\log T\rceil - 1$. If T is large, we have $M < T-1$, and hence we can write

$$\sum_{k=0}^{T-1}\exp(-\tau k)\,\frac{w_{T-k}}{w_T} = \sum_{k=0}^{M}\exp(-\tau k)\,\frac{w_{T-k}}{w_T} + \sum_{k=M+1}^{T-1}\exp(-\tau k)\,\frac{w_{T-k}}{w_T}. \tag{34}$$

Using Lemma 1, we get

$$(1-\eta_T)\sum_{k=0}^{M}\exp(-\tau k)\leqslant\sum_{k=0}^{M}\exp(-\tau k)\,\frac{w_{T-k}}{w_T}\leqslant(1+\eta_T)\sum_{k=0}^{M}\exp(-\tau k),$$

where $\eta_T = o(1)$ as $T \to \infty$. Thus

$$\lim_{T\to\infty}\sum_{k=0}^{M}\exp(-\tau k)\,\frac{w_{T-k}}{w_T} = \frac{1}{1-\mathrm{e}^{-\tau}}.$$

The last sum in (34) satisfies

$$\begin{aligned}\sum_{k=M+1}^{T-1}\exp(-\tau k)\,\frac{w_{T-k}}{w_T} &\leqslant \sum_{k=M+1}^{T-1}\exp(-\tau k)\,\left(1+\frac{C}{M^{\mu}}\right)^{k}\\ &\leqslant \sum_{k=M+1}^{\infty}\exp\left\{-(\tau - CM^{-\mu})\,k\right\}.\end{aligned}$$

This term tends to 0 as $M \to \infty$. Thus we obtain (32). Equation (33) is proved similarly.

Lemma 3. *Let ξ_i be i.i.d. $\mathscr{N}(0,1)$ random variables. Then, for any $k \in \mathbf{N}$, $N \in \mathbf{N}$ and $x > 0$,*

$$\mathbf{E}\,\xi_k^2\,\mathbf{I}\left\{\|\xi\|_k^2 \geqslant x\right\} \leqslant \left(\frac{x\mathrm{e}^2}{3+2N}\right)^{N+3/2} \exp\left(-\frac{x}{2}\right). \tag{35}$$

P r o o f. For any $0 < \lambda < \frac{1}{2}$,

$$\begin{aligned}
\mathbf{E}\xi_k^2\mathbf{I}\left\{\|\xi\|_k^2 \geqslant x\right\} &\leqslant \exp(-\lambda x)\,\mathbf{E}\xi_k^2 \exp\left\{\lambda\|\xi\|_k^2\right\} \\
&= \exp(-\lambda x)\,\mathbf{E}\xi_k^2 \exp\left\{\lambda\xi_k^2\right\}\prod_{i\neq k}\exp(\lambda\xi_1^2) \\
&= \exp(-\lambda x)(1-2\lambda)^{-3/2}(1-2\lambda)^{-N} \\
&= \exp\left(-\lambda x - \frac{3+2N}{2}\,\log(1-2\lambda)\right). \qquad (36)
\end{aligned}$$

The minimum with respect to λ of the right-hand side of (36) is attained at

$$\lambda = \frac{1}{2}\left(1 - \frac{3+2N}{x}\right).$$

Substituting this λ into (36) we get (35).

P r o o f o f T h e o r e m 1. Let M be a sufficiently large integer satisfying $N \leqslant M < \min(W^*/2, \log W^*/2)$. In this proof we denote by C the constants that do not depend on M, N, ε, and θ.

We decompose the risk of the estimator θ^* into three parts

$$\sup_{\theta\in\Theta(\alpha,L)} R^{\varepsilon}(\theta^*,\theta) \leqslant S_1 + S_2 + S_3, \tag{37}$$

where

$$S_1 = \sup_{\theta\in\Theta(\alpha,L)} \sum_{k=1}^{W^*-M} \mathbf{E}_\theta(\theta_k^* - \theta_k)^2, \qquad S_2 = \sup_{\theta\in\Theta(\alpha,L)} \sum_{k=W^*-M}^{W^*+M} \mathbf{E}_\theta(\theta_k^* - \theta_k)^2,$$

$$S_3 = \sup_{\theta\in\Theta(\alpha,L)} \sum_{k=W^*+M}^{\infty} \mathbf{E}_\theta(\theta_k^* - \theta_k)^2.$$

Consider first the term S_1. Using (32) for the subexponential sequences $w_k = r_k$ and $w_k = kr_k$, and Lemma 1 for $w_k = r_k$, we have

$$\begin{aligned}
S_1 &= \sup_{\theta\in\Theta(\alpha,L)} \sum_{k=1}^{W^*-M} \exp(\rho k)\, r_k \mathbf{E}_\theta\left[y_k\,\mathbf{I}\left\{\|y\|_k^2 < 2\,\varepsilon^2\rho' k\right\} - \varepsilon\xi_k\right]^2 \\
&\leqslant 2\,\varepsilon^2 \sum_{k=1}^{W^*-M} \exp(\rho k)\, r_k
\end{aligned}$$

$$
\begin{aligned}
&+ 2 \sup_{\theta\in\Theta(\alpha,L)} \sum_{k=1}^{W^*-M} \exp(\rho k)\, r_k \mathbf{E}_\theta y_k^2\, \mathbf{I}\left\{\|y\|_k^2 < 2\varepsilon^2\rho' k\right\} \\
&\leqslant 2\varepsilon^2 \sum_{k=1}^{W^*-M} \exp(\rho k)\, r_k + 4\varepsilon^2 \sum_{k=1}^{W^*-M} \exp(\rho k)\, r_k k\rho' \\
&\leqslant C\rho\varepsilon^2 \exp(\rho W^*)\, r_{W^*} W^* \exp(-\rho M) \leqslant C\,\psi_\varepsilon(\alpha, L)\exp(-\rho M). \quad (38)
\end{aligned}
$$

Next, we bound from above the term S_2. This is the main term in (37) and we will analyze it using a renormalization argument. We begin with some simple remarks. Denote

$$
\Theta_M = \left\{\theta = (\theta_1, \theta_2, \dots)\colon \sum_{k=W^*-M-N}^{W^*+M+N} \exp(\gamma k)\, q_k \theta_k^2 \leqslant L\right\}.
$$

Clearly, $\Theta(\alpha, L) \subset \Theta_M$. Now we change the variables from θ_k to ν_k by setting, for $k \geqslant 1 - W^*$,

$$
\nu_k = \frac{\theta_{k+W^*}}{\varepsilon\sqrt{2\rho W^* r_{k+W^*}\exp(\rho(k+W^*))}} = \theta_{k+W^*}\left(\frac{A^* r_{W^*}}{\psi_\varepsilon(\alpha, L)\,\exp(\rho k)\, r_{k+W^*}}\right)^{1/2},
$$

and let ν_k^* be derived from θ_k^* by the same transformation. If $\theta \in \Theta_M$, the sequence $\nu = \{\nu_k\}$ belongs to the set

$$
\Xi_M = \left\{\nu \in \Xi\colon \sum_{k=-M-N}^{M+N} \exp(\gamma k)\, r_{k+W^*} q_{k+W^*} \nu_k^2 \leqslant \frac{L\exp(-\gamma W^*)}{2\varepsilon^2\rho W^*}\right\}
$$

where Ξ denotes the set of all sequences of the form $\nu = (\nu_{1-W^*}, \dots, \nu_0, \nu_1, \dots)$. Now, in view of Lemma 1, applied to the subexponential sequences $w_k = r_k$ and $w_k = r_k q_k$, there exists $\eta = \eta_\varepsilon$ depending only on W^*, such that $\eta \to 0$, as $\varepsilon \to 0$, and

$$
\max_{|k|\leqslant M} r_{k+W^*} \leqslant (1+\eta)\, r_{W^*}, \quad (39)
$$

$$
\min_{|k|\leqslant M+N} r_{k+W^*} q_{k+W^*} \geqslant (1-\eta)\, r_{W^*} q_{W^*}. \quad (40)
$$

Fix $0 < \delta < 1$, and assume that ε is small enough to have simultaneously $\eta < \delta$ and $(1-\eta)^{-1} < 1 + \delta$. Then (40) implies that $\min_{|k|\leqslant M+N} r_{k+W^*} q_{k+W^*} \geqslant r_{W^*} q_{W^*}/(1+\delta)$, and therefore

$$
\Xi_M \subset \Xi_M^\delta = \left\{\nu \in \Xi\colon \sum_{k=-M-N}^{M+N} \exp(\gamma k)\, \nu_k^2 \leqslant E^*(1+\delta)\right\}.
$$

Furthermore, (39) guarantees that $\max_{|k|\leqslant M} r_{k+W^*} \leqslant (1+\delta)\, r_{W^*}$. These remarks imply that, for ε small enough,

$$
S_2 \leqslant \sup_{\theta\in\Theta_M} \sum_{k=W^*-M}^{W^*+M} \mathbf{E}_\theta(\theta_k^* - \theta_k)^2
$$

$$
\begin{aligned}
&\leqslant \frac{\psi_\varepsilon(\alpha, L)}{A^*} \sup_{\nu \in \Xi_M^\delta} \sum_{k=-M}^{M} \exp(\rho k) \frac{r_{k+W^*}}{r_{W^*}} \mathbf{E}_\theta(\nu_k^* - \nu_k)^2 \\
&\leqslant \frac{(1+\delta)\, \psi_\varepsilon(\alpha, L)}{A^*} \sup_{\nu \in \Xi_M^\delta} \sum_{k=-M}^{M} \exp(\rho k)\, \mathbf{E}_\theta(\nu_k^* - \nu_k)^2. \qquad (41)
\end{aligned}
$$

Using the inequality $(x+y)^2 \leqslant (1+\delta)\, x^2 + (1+\delta^{-1})\, y^2$ for any $x, y \in \mathbf{R}$, we get

$$
\begin{aligned}
\mathbf{E}_\theta(\nu_k^* - \nu_k)^2 \;=\; &\mathbf{E}\Bigg[\left(\nu_k + \frac{\xi_{k+W^*}}{\sqrt{2\rho W^*}}\right) \\
&\times \mathbf{I}\left\{\sum_{l=k-N}^{k+N}\left(\nu_l + \frac{\xi_{l+W^*}}{\sqrt{2\rho W^*}}\right)^2 \geqslant \frac{\rho'}{\rho}\left(1 + \frac{k}{W^*}\right)\right\} - \nu_k\Bigg]^2 \\
\leqslant\; &(1+\delta)\, \nu_k^2 \mathbf{P}\left\{\sum_{l=k-N}^{k+N}\left(\nu_l + \frac{\xi_{l+W^*}}{\sqrt{2\rho W^*}}\right)^2 < \frac{\rho'}{\rho}\left(1 + \frac{k}{W^*}\right)\right\} \\
&+ (1+\delta^{-1})(2\rho W^*)^{-1}. \qquad (42)
\end{aligned}
$$

Next, using the inequality $(x+y)^2 \geqslant (1-\delta)x^2 + (1-\delta^{-1})y^2$ for any $x, y \in \mathbf{R}$, one obtains

$$
\sum_{l=k-N}^{k+N}\left(\nu_l + \frac{\xi_{l+W^*}}{\sqrt{2\rho W^*}}\right)^2 \geqslant (1-\delta) \sum_{l=k-N}^{k+N} \nu_l^2 - \frac{1-\delta}{2\rho W^* \delta} \sum_{l=k-N}^{k+N} \xi_{l+W^*}^2
$$

and, therefore,

$$
\begin{aligned}
&\mathbf{P}\left\{\sum_{l=k-N}^{k+N}\left(\nu_l + \frac{\xi_{l+W^*}}{\sqrt{2\rho W^*}}\right)^2 < \frac{\rho'}{\rho}\left(1 + \frac{k}{W^*}\right)\right\} \\
&\leqslant \mathbf{P}\left\{(1-\delta) \sum_{l=k-N}^{k+N} \nu_l^2 - \frac{1-\delta}{2\rho W^* \delta} \sum_{l=k-N}^{k+N} \xi_{l+W^*}^2 < \frac{\rho'}{\rho}\left(1 + \frac{k}{W^*}\right)\right\} \\
&\leqslant \mathbf{I}\left\{\sum_{l=k-N}^{k+N} \nu_l^2 \leqslant V\right\} + \mathbf{P}\left\{\frac{1}{4\delta\rho W^*} \sum_{l=k-N}^{k+N} \xi_{l+W^*}^2 > \delta\right\}, \qquad (43)
\end{aligned}
$$

where

$$
V = 2\delta + \frac{1}{1-\delta}\left(1 + \frac{M}{W^*}\right)\frac{\rho'}{\rho}.
$$

By Markov's inequality

$$
\mathbf{P}\left\{\frac{1}{4\delta\rho W^*} \sum_{l=k-N}^{k+N} \xi_{l+W^*}^2 > \delta\right\} \leqslant \frac{2N+1}{4\delta^2 \rho W^*}. \qquad (44)
$$

Define

$$
A(N, M) = \sup_{\nu \in \Xi_M^\delta} \sum_{k=-M}^{M} \exp(\rho k)\, \nu_k^2\, \mathbf{I}\left\{\sum_{l=k-N}^{k+N} \nu_l^2 \leqslant V\right\}.
$$

Combining (42)–(44) one obtains

$$\begin{aligned}\sup_{\nu\in\Xi_M^\delta}\sum_{k=-M}^{M}\exp(\rho k)\,\mathbf{E}_\theta(\nu_k^*-\nu_k)^2 &\leqslant (1+\delta)\\ &+\left[A(N,M)+\frac{2N+1}{4\delta^2\rho W^*}\sup_{\nu\in\Xi_M^\delta}\sum_{k=-M}^{M}\exp(\rho k)\,\nu_k^2\right]+\frac{C\exp(\rho M)}{W^*}\\ &\leqslant (1+\delta)\,A(N,M)+\frac{C}{W^*}\,(N+\exp(\rho M)),\end{aligned} \tag{45}$$

where for the last inequality we used that, since $\gamma>\rho$,

$$\sup_{\nu\in\Xi_M^\delta}\sum_{k=-M}^{M}\exp(\rho k)\,\nu_k^2\leqslant\sup_{\nu\in\Xi_M^\delta}\sum_{k=-M}^{M}\exp(\gamma k)\,\nu_k^2\leqslant E^*(1+\delta).$$

From (41) and (45) we get that for any $0<\delta<1$ and for all ε small enough

$$S_2\leqslant\frac{\psi_\varepsilon(\alpha,L)}{A^*}\left[(1+\delta)^2A(N,M)+\frac{C(N+\exp(\rho M))}{W^*}\right]. \tag{46}$$

Our next goal is to show that $A(N,M)$ is close to A^*. We will proceed in steps. The first step is to remark that

$$\begin{aligned}A(N,M)&\leqslant\sup_{\nu\in\Xi_M^\delta}\sum_{k=-M}^{M}\exp(\rho k)\,\nu_k^2\,\mathbf{I}\left\{\sum_{l=k-N}^{k+N}\nu_l^2\leqslant V(1+\delta)\right\}\\ &\leqslant V\left[\sup_{\nu\in\Xi_*}\sum_{k=-M}^{M}\exp(\rho k)\,\nu_k^2\,\mathbf{I}\left\{\sum_{l=k-N}^{k+N}\nu_l^2\leqslant 1\right\}\right],\end{aligned} \tag{47}$$

where

$$\Xi_*=\left\{\nu\in\Xi\colon\sum_{k=-M-N}^{M+N}\exp(\gamma k)\,\nu_k^2\leqslant E^*\right\}.$$

In fact, to get (47), introduce the sequence $\nu'\in\Xi$ such that $\nu_k=\nu_k'\sqrt{V(1+\delta)}$, observe that $V>1$ and use the embedding

$$\begin{aligned}&\left\{\nu'\in\Xi\colon V\sum_{k=-M-N}^{M+N}\exp(\gamma k)(\nu_k')^2\leqslant E^*\right\}\\ &\subseteq\left\{\nu'\in\Xi\colon\sum_{k=-M-N}^{M+N}\exp(\gamma k)(\nu_k')^2\leqslant E^*\right\}.\end{aligned}$$

Our next step is to find an upper bound for the expression in square brackets in (47). We can write assuming without loss of generality that N is even,

$$\sup_{\nu\in\Xi_*}\sum_{k=-M-N}^{M+N}\exp(\rho k)\,\nu_k^2\,\mathbf{I}\left\{\sum_{l=k-N}^{k+N}\nu_l^2\leqslant 1\right\}\leqslant A_1+A_2+A_3 \tag{48}$$

with

$$\begin{aligned} A_1 &= \sup_{\nu\in\Xi_*} \sum_{k=-M-N}^{-N/2-1} \exp(\rho k)\,\nu_k^2\,\mathbf{I}\bigg\{\sum_{l=k-N}^{k+N}\nu_l^2 \leqslant 1\bigg\} \leqslant \sum_{k=-\infty}^{-N/2-1} \mathrm{e}^{\rho k} \\ &\leqslant C\exp\Big(-\frac{\rho N}{2}\Big), \\ A_2 &= \sup_{\nu\in\Xi_*} \sum_{k=N/2+1}^{M+N} \exp(\rho k)\,\nu_k^2\,\mathbf{I}\bigg\{\sum_{l=k-N}^{k+N}\nu_l^2 \leqslant 1\bigg\} \leqslant \sup_{\nu\in\Xi_*}\sum_{k=N/2+1}^{M+N} \mathrm{e}^{\rho k}\nu_k^2 \\ &\leqslant C\exp\Big(-\frac{\alpha N}{2}\Big), \end{aligned}$$

and

$$\begin{aligned} A_3 &= \sup_{\nu\in\Xi_*} \sum_{k=-N/2}^{N/2} \exp(\rho k)\,\nu_k^2\,\mathbf{I}\bigg\{\sum_{l=k-N}^{k+N}\nu_l^2 \leqslant 1\bigg\} \\ &\leqslant \sup_{\nu\in\Xi_*} \sum_{k=-N/2}^{N/2} \exp(\rho k)\,\nu_k^2\,\mathbf{I}\bigg\{\sum_{l=-N/2}^{N/2}\nu_l^2 \leqslant 1\bigg\} \leqslant \sup_{\nu\in\Xi'}\sum_{k=-N/2}^{N/2} \exp(\rho k)\,\nu_k^2, \end{aligned}$$

where

$$\Xi' = \bigg\{\nu\in\Xi\colon \sum_{k=-M-N}^{M+N}\exp(\gamma k)\,\nu_k^2 \leqslant E^*,\ \sum_{k=-N/2}^{N/2}\nu_k^2\leqslant 1\bigg\}.$$

Substitution of the inequalities for A_1, A_2, and A_3 into (48) yields

$$\begin{aligned} &\sup_{\nu\in\Xi_*} \sum_{k=-M-N}^{M+N} \exp(\rho k)\,\nu_k^2\,\mathbf{I}\bigg\{\sum_{l=k-N}^{k+N}\nu_l^2 \leqslant 1\bigg\} \\ &\qquad \leqslant \sup_{\nu\in\Xi'}\sum_{k=-N/2}^{N/2}\exp(\rho k)\,\nu_k^2 + C\exp\Big(-\frac{N}{2}\min(\alpha,\rho)\Big). \end{aligned} \tag{49}$$

Next, introducing the set $\Theta_\infty^{N/2}(\alpha,L) = \{\nu\in\Theta_\infty(\alpha,L)\colon \nu_k=0,\ \text{for } |k|>N/2\}$, we note that

$$\begin{aligned} \sup_{\nu\in\Xi'}\sum_{k=-N/2}^{N/2}\exp(\rho k)\,\nu_k^2 &= \sup_{\nu\in\Theta_\infty^{N/2}(\alpha,L)}\sum_{k=-N/2}^{N/2}\exp(\rho k)\,\nu_k^2 \\ &\leqslant \sup_{\nu\in\Theta_\infty(\alpha,L)}\sum_{k=-N/2}^{N/2}\exp(\rho k)\,\nu_k^2 \leqslant A^*. \end{aligned} \tag{50}$$

In fact, the equality in (50) follows from the fact that taking $\nu_k\neq 0$ for $|k|>N/2$ does not increase the sum $\sum_{k=-N/2}^{N/2}\exp(\rho k)\,\nu_k^2$, and thus the supremum of this sum over Ξ' is attained on the sequences ν with $\nu_k = 0$ for $|k| > N/2$.

From (47), (49)–(50) we get $A(N, M) \leqslant V[A^* + C \exp(-N \min(\alpha, \rho)/2)]$. This together with (46) and fact that A^* is bounded from below uniformly in ε entail

$$S_2 \leqslant \psi_\varepsilon(\alpha, L) \Big[V(1+\delta)^2 + C \exp\Big(-\frac{N}{2} \min(\alpha, \rho)\Big) + \frac{C}{W^*} (N + \exp(\rho M)) \Big]. \tag{51}$$

Finally we bound from above the term S_3. Using (3) and (10) we find

$$S_3 \leqslant 2 \sup_{\theta \in \Theta(\alpha, L)} \sum_{k=W^*+M}^{\infty} \theta_k^2 + 2\varepsilon^2 \sup_{\theta \in \Theta(\alpha, L)} \sum_{k=W^*+M}^{\infty} \exp(\rho k)\, r_k \mathbf{E}\, \xi_k^2\, \mathbf{I}\{\|y\|_k^2 \geqslant 2\varepsilon^2 \rho' k\}. \tag{52}$$

The first term in the right-hand side satisfies

$$\begin{aligned} \sup_{\theta \in \Theta(\alpha, L)} \sum_{k=W^*+M}^{\infty} \theta_k^2 &= \sup_{\theta \in \Theta(\alpha, L)} \sum_{k=W^*+M}^{\infty} \exp(-\alpha k)\, q_k^{-1}\, q_k \theta_k^2 \exp(\alpha k) \\ &\leqslant \sup_{\theta \in \Theta(\alpha, L)} \sum_{k=W^*+M}^{\infty} \exp(-\alpha k)\, q_k^{-1} \sum_{k=1}^{\infty} q_k \exp(\alpha k)\, \theta_k^2 \\ &\leqslant C \exp[-\alpha(W^* + M)]\, q_{W^*+M}^{-1} \\ &\leqslant C \exp[-\alpha M]\, \psi_\varepsilon(\alpha, L), \end{aligned} \tag{53}$$

where to get the last two inequalities we applied Lemma 2 and then Lemma 1 for the subexponential sequence $w_k = q_k^{-1}$.

Consider the sequence $\theta' = (\theta'_1, \theta'_2, \ldots)$ with $\theta'_k = b_k \theta_k$. For $k \geqslant W^* + M$

$$\begin{aligned} \sup_{\theta \in \Theta(\alpha, L)} \|\theta'\|_k^2 &= \sup_{\theta \in \Theta(\alpha, L)} \sum_{l=k-N}^{k+N} \theta_l^2 r_l^{-1} \exp(-\rho l) \\ &\leqslant \sup_{\theta \in \Theta(\alpha, L)} \sum_{l=k-N}^{k+N} q_l \exp(\alpha l)\, \theta_l^2 \sum_{l=k-N}^{k+N} r_l^{-1} q_l^{-1} \exp(-\gamma l) \\ &\leqslant L \sum_{l=k-N}^{k+N} r_l^{-1} q_l^{-1} \exp(-\gamma l) \leqslant C r_{k-N}^{-1} q_{k-N}^{-1} \exp(-\gamma(k-N)), \end{aligned}$$

where we used Lemma 2 for the subexponential sequence $w_l = r_l^{-1} q_l^{-1}$. Applying to the last expression of the previous display successively Lemma 1 and then the fact that $L \exp(-\gamma W^*) < 2\varepsilon^2 \rho W^* r_{W^*} q_{W^*}$, we get

$$\begin{aligned} \sup_{\theta \in \Theta(\alpha, L)} \|\theta'\|_k^2 &\leqslant C r_{W^*}^{-1} q_{W^*}^{-1} \exp(-\gamma(k-N)) \\ &\leqslant C\varepsilon^2 W^* \exp(-\gamma(M-N)) \leqslant C\varepsilon^2 k \exp(-\gamma(M-N)). \end{aligned} \tag{54}$$

Now we bound the last term in (52). By Lemma 3, for any $k \geqslant W^* + M$ we have

$$\begin{aligned}
\mathbf{E}\xi_k^2 \mathbf{I}\left\{\|\theta' + \varepsilon\xi\|_k^2 \geqslant 2\varepsilon^2\rho' k\right\} &\leqslant \mathbf{E}\,\xi_k^2\,\mathbf{I}\left\{\|\xi\|_k \geqslant \sqrt{2\rho' k} - \left\|\frac{\theta'}{\varepsilon}\right\|_k\right\} \\
&\leqslant \left(\frac{2\rho' k \mathrm{e}^2}{3+2N}\right)^{N+3/2} \exp\left\{-\frac{1}{2}\left[\sqrt{2\rho' k} - \left\|\frac{\theta'}{\varepsilon}\right\|_k\right]^2\right\} \\
&\leqslant \left(\frac{2\rho' k \mathrm{e}^2}{3+2N}\right)^{N+3/2} \exp\left\{-\rho' k + \sqrt{2\rho' k}\left\|\frac{\theta'}{\varepsilon}\right\|_k\right\}. \qquad (55)
\end{aligned}$$

In the rest of the proof we set

$$M = 2\left\lceil \sqrt{\log[\log(\varepsilon^{-1}) \vee 1]}\right\rceil. \qquad (56)$$

For ε small enough this choice of M satisfies the assumptions on M imposed above, since $W^* \asymp \log(1/\varepsilon)$. Since $M \to \infty$ as $\varepsilon \to 0$, for any small constant $c > 0$ there exists ε_0 such that for $\varepsilon < \varepsilon_0$ we have $\exp[-\gamma(M-N)] \leqslant c(\rho'-\rho)^2$. Then for $\varepsilon < \varepsilon_0$, in view of (54) we have $-(\rho'-\rho)k + \sqrt{2\rho' k}\|\theta'/\varepsilon\|_k \leqslant -(\rho'-\rho)k/2$ and, therefore,

$$\begin{aligned}
&\sup_{\theta\in\Theta(\alpha,L)} \sum_{k=W^*+M}^{\infty} \exp(\rho k)\, r_k \mathbf{E}\xi_k^2\,\mathbf{I}\left\{\|\theta' + \varepsilon\xi\|_k^2 \geqslant 2\,\varepsilon^2\rho' k\right\} \\
&\leqslant \sup_{\theta\in\Theta(\alpha,L)} \sum_{k=W^*+M}^{\infty} \left(\frac{2\rho' k \mathrm{e}^2}{3+2N}\right)^{N+3/2} r_k \exp\left\{-(\rho'-\rho)\,k + \sqrt{2\rho' k}\left\|\frac{\theta'}{\varepsilon}\right\|_k\right\} \\
&\leqslant \sum_{k=W^*+M}^{\infty} \left(\frac{2\rho' k \mathrm{e}^2}{3+2N}\right)^{N+3/2} r_k \exp\left\{-(\rho'-\rho)\,\frac{k}{2}\right\} \\
&\leqslant C r_{W^*}\left(\frac{2\rho'(W^*+M)\mathrm{e}^2}{3+2N}\right)^{N+3/2} (\rho'-\rho)^{-1} \exp\left\{-(\rho'-\rho)\,\frac{W^*}{2}\right\},
\end{aligned}$$

where the last inequality follows from (33) of Lemma 2 with the subexponential sequence $w_k = k^{N+3/2} r_k$ and from Lemma 1 with $w_k = r_k$. This inequality and (52), (53) yield

$$\begin{aligned}
S_3 &\leqslant C[\exp(-\alpha M)\,\psi_\varepsilon(\alpha, L) + (\rho'-\rho)^{-1}\varepsilon^2 r_{W^*}] \\
&\leqslant C\psi_\varepsilon(\alpha, L)\,\left[\exp(-\alpha M) + ((\rho'-\rho)\,W^*)^{-1}\right].
\end{aligned}$$

Combining this result with (37), (38), and (51) we find

$$\begin{aligned}
\sup_{\theta\in\Theta(\alpha,L)} \frac{R^\varepsilon(\theta^*,\theta)}{\psi_\varepsilon(\alpha,L)} &\leqslant V(1+\delta)^2 \\
&\quad + C\Big[\exp(-\alpha M) + \exp(-\rho M) + \exp\left(-\frac{N}{2}\min(\alpha,\rho)\right) \\
&\quad + \frac{N + \exp(\rho M) + (\rho'-\rho)^{-1}}{W^*}\Big].
\end{aligned}$$

It remains to take limits as $\varepsilon \to 0$, and then as $\delta \to 0$, using the definition of M in (56) and the definition of V. This completes the proof of (26).

P r o o f o f T h e o r e m 2. We follow the lines of the proof of Theorem 1 with $M = 2N$ (cf. (56)). The argument preceding (56) is true for any fixed N and $\rho' > \rho$, and it remains intact. Inspection of the proof of Theorem 1 after (56) shows that the choice of N and ρ' defined in Theorem 2 is sufficient to get (27).

4.2. Proof of Theorem 3.

For an integer M, consider the set

$$\Theta_\infty^M(\alpha, L) = \{\theta \in \Theta_\infty(\alpha, L) \colon \theta_k = 0, \text{ for } |k| > M\}$$

and define

$$A_M(\alpha, L) = \sup_{\theta \in \Theta_\infty^M(\alpha, L)} \sum_{k=-M}^{M} \exp(\rho k)\, \theta_k^2.$$

Lemma 4. *For any* $\alpha > 0$, $L > 0$,

$$\lim_{M \to \infty} A_M(\alpha, L) = A^*(\alpha, L). \tag{57}$$

P r o o f. Fix $\alpha > 0$, $L > 0$, and omit for brevity the indication of α and L in brackets for $A_M, A^*, \Theta_\infty, \Theta_\infty^M, E^*$. Obviously, $A_M \leqslant A^*$, and we have to show only that

$$\liminf_{M \to \infty} A_M \geqslant A^*. \tag{58}$$

We first prove that

$$A_M \geqslant Q \sup_{\theta \in \Theta_\infty} \sum_{k=-M}^{M} \exp(\rho k)\, \theta_k^2, \tag{59}$$

where $Q = E^*/(E^* + \exp(-\gamma M))$. To do this, it suffices to show that for any $\theta \in \Theta_\infty$ there exists $\theta' \in \Theta_\infty^M$ such that

$$\sum_{k=-M}^{M} \exp(\rho k)(\theta_k')^2 \geqslant Q \sum_{k=-M}^{M} \exp(\rho k)\, \theta_k^2. \tag{60}$$

If $\theta \in \Theta_\infty^M$ this inequality is obvious (one takes $\theta' = \theta$). If $\theta \in \Theta_\infty \setminus \Theta_\infty^M$, we have $S = \sum_{|k|>M} \theta_k^2 > 0$. Also, $S \leqslant 1$ since $\theta \in \Theta_\infty$. Define θ' by

$$\begin{aligned} \theta_k' &= Q^{1/2}\theta_k, \qquad -M < k \leqslant M, \\ \theta_{-M}' &= Q^{1/2}\sqrt{\theta_{-M}^2 + S}, \quad \theta_k' = 0, \qquad |k| > M. \end{aligned}$$

Clearly, $\theta' \in \Theta_\infty^M$ since $\sum_{k=-M}^{M} (\theta_k')^2 \leqslant \sum_{k=-M}^{M} \theta_k^2 \leqslant 1$ and

$$\begin{aligned}\sum_{k=-M}^{M} \exp(\gamma k)(\theta_k')^2 &= Q\left[\sum_{k=-M}^{M} \exp(\gamma k)\,\theta_k^2 + S\exp(-\gamma M)\right]\\ &\leqslant Q\left[\exp(-\gamma M)\sum_{k\leqslant -M}\theta_k^2 + \sum_{k=-M+1}^{\infty}\exp(\gamma k)\,\theta_k^2\right]\\ &\leqslant Q\left[\exp(-\gamma M) + \sum_{k=-\infty}^{\infty}\exp(\gamma k)\,\theta_k^2\right] \leqslant E^*.\end{aligned}$$

On the other hand,

$$\begin{aligned}\sum_{k=-M}^{M} \exp(\rho k)(\theta_k')^2 &= Q\left[\sum_{k=-M}^{M} \exp(\rho k)\,\theta_k^2 + S\exp(-\rho M)\right]\\ &\geqslant Q\sum_{k=-M}^{M}\exp(\rho k)\,\theta_k^2.\end{aligned}$$

This proves (60) and therefore (59). To finish the proof of (58) it remains to combine (59) with the following inequality:

$$\begin{aligned}\sup_{\theta\in\Theta_\infty}\sum_{k=-M}^{M}\exp(\rho k)\,\theta_k^2 &\geqslant A^* - \sup_{\theta\in\Theta_\infty}\left(\sum_{k<-M}\exp(\rho k)\,\theta_k^2 + \sum_{k>M}\exp(\rho k)\,\theta_k^2\right)\\ &\geqslant A^* - E^*\exp(-\alpha M) - \exp(-\rho M).\end{aligned}$$

Lemma 5. *Under the assumptions of Theorem* 3, *for any* $0 < \alpha_0 < \alpha' \leqslant \alpha$ *and any* $L_0 > 0$, $L > 0$ *we have*

$$\liminf_{\varepsilon\to\infty}\inf_{\hat\theta}\max\left\{\sup_{\theta\in\Theta(\alpha_0,L_0)}\frac{R^\varepsilon(\hat\theta,\theta)}{\psi_\varepsilon(\alpha_0,L_0)},\ \sup_{\theta\in\Theta(\alpha,L)}\frac{R^\varepsilon(\hat\theta,\theta)}{\psi_\varepsilon(\alpha',L)}\right\} \geqslant 1 - \frac{\gamma_0}{\gamma'}, \tag{61}$$

where $\inf_{\hat\theta}$ *denotes the infimum over all estimators and* $\gamma_0 = \alpha_0 + \rho$, $\gamma' = \alpha' + \rho$.

P r o o f. We will write for brevity $W_0 = W^*(\alpha_0, L_0)$, $W' = W^*(\alpha', L)$. Let M be an integer satisfying $1 \leqslant M < W_0$, and let $\delta \in (0,1)$ be such that $1 - W'(1+\delta)/W_0 > 0$. Such a choice of δ is possible for all ε small enough since under the assumption $q_k \equiv 1$ we have, in view of (21),

$$\lim_{\varepsilon\to 0}\frac{W'}{W_0} = \frac{\gamma_0}{\gamma'} < 1. \tag{62}$$

For $\widetilde{L} = L_0[1 - W'(1+\delta)/W_0]$ set

$$\Theta_{M,0} = \left\{\theta = (\theta_1, \theta_2, \ldots)\colon \sum_{k=W_0-M}^{W_0+M}\exp(\alpha_0 k)\,\theta_k^2 \leqslant \widetilde{L} \text{ and } \theta_k = 0,\ |k - W_0| > M\right\}.$$

Clearly, $\Theta_{M,0} \subset \Theta(\alpha_0, L_0)$, and, therefore,

$$r^\varepsilon \stackrel{\text{def}}{=} \inf_{\hat\theta} \max\left\{ \sup_{\theta\in\Theta(\alpha_0,L_0)} \frac{R^\varepsilon(\hat\theta,\theta)}{\psi_\varepsilon(\alpha_0,L_0)}, \quad \sup_{\theta\in\Theta(\alpha,L)} \frac{R^\varepsilon(\hat\theta,\theta)}{\psi_\varepsilon(\alpha',L)} \right\}$$

$$\geqslant \sup_{\theta\in\Theta_{M,0}\setminus\{\mathbf{0}\}} \inf_{\hat\theta} \max\left\{ \frac{R^\varepsilon(\hat\theta,\theta)}{\psi_\varepsilon(\alpha_0,L_0)}, \frac{R^\varepsilon(\hat\theta,\mathbf{0})}{\psi_\varepsilon(\alpha',L)} \right\}, \tag{63}$$

where $\mathbf{0}$ denotes the sequence θ with all the elements equal to 0. To handle the last expression, we use again a renormalization. Change the variables from θ_k to ν_k by setting, for $k \geqslant 1 - W_0$,

$$\nu_k = \frac{b_{k+W_0}\theta_{k+W_0}}{\varepsilon\sqrt{2\rho W_0}} = \frac{\theta_{k+W_0}}{\varepsilon\sqrt{2\rho W_0 \exp(\rho(k+W_0))}},$$

and let $\widehat\nu_k$ be obtained from $\hat\theta_k$ by the same transformation, thus defining a sequence $\widehat\nu \in \Xi$. We will also write $\mathbf{P}_\nu$, $\mathbf{E}_\nu$ instead of $\mathbf{P}_\theta$ and $\mathbf{E}_\theta$, respectively.

Clearly, $\theta \in \Theta_{M,0}$ if and only if $\nu = \{\nu_k\}$ belongs to the set

$$\Xi_{M,0} = \left\{ \nu \in \Xi\colon \sum_{k=-M}^{M} \exp(\gamma_0 k)\, \nu_k^2 \leqslant \widetilde E \text{ and } \nu_k = 0,\ |k| > M \right\},$$

where $\widetilde E = E^*(\alpha_0, L_0)[1 - W'(1+\delta)/W_0]$. With this notation,

$$\frac{R^\varepsilon(\hat\theta,\theta)}{\psi_\varepsilon(\alpha_0,L_0)} = \frac{1}{A^*(\alpha_0,L_0)} \left[\sum_{k=-M}^{M} \exp(\rho k)\, \mathbf{E}_\nu(\widehat\nu_k - \nu_k)^2 + \sum_{|k|>M} \exp(\rho k)\, \mathbf{E}_\nu \widehat\nu_k^2 \right]$$

$$\geqslant \frac{1}{A^*(\alpha_0,L_0)} \sum_{k=-M}^{M} \exp(\rho k)\, \mathbf{E}_\nu(\widehat\nu_k - \nu_k)^2$$

which entails, together with (63), that

$$r^\varepsilon \geqslant \sup_{\nu\in\Xi_{M,0}\setminus\{\mathbf{0}\}} \left[\frac{(1-\delta_0)^{-2}}{A^*(\alpha_0,L_0)} \left(\sum_{k=-M}^{M} \exp(\rho k)\, \nu_k^2 \right) D_\nu^\varepsilon \right], \tag{64}$$

where $0 < \delta_0 < 1$,

$$D_\nu^\varepsilon = \inf_{\widehat\nu} \max\left\{ \mathbf{E}_\nu d_\nu^2(\widehat\nu, \nu),\ \lambda^\varepsilon \mathbf{E}_{\mathbf{0}} d_\nu^2(\widehat\nu, \mathbf{0}) \right\}, \qquad \lambda^\varepsilon = \frac{\psi_\varepsilon(\alpha_0,L_0)}{\psi_\varepsilon(\alpha',L)},$$

and $d_\nu(\nu^{(1)}, \nu^{(2)})$, for a fixed $\nu \in \Xi \setminus \{\mathbf{0}\}$, denotes the distance between two sequences $\nu^{(1)} \in \Xi$ and $\nu^{(2)} \in \Xi$ defined by

$$d_\nu^2(\nu^{(1)}, \nu^{(2)}) = (1-\delta_0)^2 \frac{\sum_{k=-M}^{M} \exp(\rho k)(\nu_k^{(1)} - \nu_k^{(2)})^2}{\sum_{k=-M}^{M} \exp(\rho k)\, \nu_k^2}.$$

In particular, $d_\nu(\nu, \mathbf{0}) = 1 - \delta_0$, which allows us to apply Theorem 6 of [18] resulting in

$$D_\nu^\varepsilon \geqslant \frac{\tau\lambda^\varepsilon\delta_0^2(1-2\delta_0)^2}{(1-2\delta_0)^2 + \tau\lambda^\varepsilon\delta_0^2}\,\mathbf{P}_\nu\Big(\frac{d\mathbf{P_0}}{d\mathbf{P}_\nu} \geqslant \tau\Big) \tag{65}$$

for any $\tau > 0$. Now, $\mathbf{P}_\nu$ is the Gaussian measure corresponding to the observations $Y_k = \nu_k + \xi_{k+W_0}/\sqrt{2\rho W_0}$ (here Y_k is the value obtained from y_k by the same transformation as ν_k is obtained from θ_k). Thus,

$$\mathbf{P}_\nu\Big(\frac{d\mathbf{P_0}}{d\mathbf{P}_\nu} \geqslant \tau\Big) = \mathbf{P}\Big\{\exp(\xi\sqrt{2\rho W_0}\|\nu\| - \rho W_0\|\nu\|^2) \geqslant \tau\Big\},$$

where $\|\nu\| = (\sum_{k=-M}^M \nu_k^2)^{1/2}$ and ξ is the standard Gaussian random variable. Set $\tau = \exp(\rho(W' - W_0) + \delta\rho W'/2)$. Then, for $\|\nu\|^2 \leqslant [1 - W'(1+\delta)/W_0]$ we get

$$\begin{aligned}
\mathbf{P}_\nu\Big(\frac{d\mathbf{P_0}}{d\mathbf{P}_\nu} \geqslant \tau\Big) &\geqslant \mathbf{P}\Big\{\xi \geqslant \frac{1}{\sqrt{2\rho W_0}\|\nu\|}(\log\tau + \rho W_0\|\nu\|^2)\Big\} \\
&\geqslant \mathbf{P}\Big\{\xi \geqslant -\frac{\delta\rho W'}{2\sqrt{2\rho W_0}\|\nu\|}\Big\} \\
&\geqslant \mathbf{P}\Big\{\xi \geqslant -\frac{\delta\rho W'}{2\sqrt{2\rho W_0}}\Big[1 - \frac{W'}{W_0}(1+\delta)\Big]^{-1/2}\Big\} \\
&\stackrel{\text{def}}{=} p_\varepsilon = 1 + o(1) \qquad \text{as} \quad \varepsilon \to 0,
\end{aligned} \tag{66}$$

in view of (62) and since $W' \to \infty$ as $\varepsilon \to 0$. Hence, introducing the set

$$\Xi'_{M,0} = \Big\{\nu \in \Xi_{M,0}\colon \|\nu\|^2 \leqslant \Big[1 - \frac{W'}{W_0}(1+\delta)\Big]\Big\}$$

and using (64)–(66) we get

$$r^\varepsilon \geqslant \frac{p_\varepsilon\tau\lambda^\varepsilon\delta_0^2(1-2\delta_0)^2(1-\delta_0)^{-2}}{(1-2\delta_0)^2 + \tau\lambda^\varepsilon\delta_0^2}\,(A^*(\alpha_0, L_0))^{-1} \sup_{\nu\in\Xi'_{M,0}\setminus\{\mathbf{0}\}} \sum_{k=-M}^{M} \exp(\rho k)\,\nu_k^2. \tag{67}$$

Next, note that

$$\begin{aligned}
&\sup_{\nu\in\Xi'_{M,0}\setminus\{\mathbf{0}\}} \sum_{k=-M}^{M} \exp(\rho k)\,\nu_k^2 \\
&\quad = \Big[1 - \frac{W'}{W_0}(1+\delta)\Big] \sup_{\nu\in\Theta_\infty^M(\alpha_0,L_0)\setminus\{\mathbf{0}\}} \sum_{k=-M}^{M} \exp(\rho k)\,\nu_k^2 \\
&\quad = \Big[1 - \frac{W'}{W_0}(1+\delta)\Big] A_M(\alpha_0, L_0).
\end{aligned} \tag{68}$$

In fact, the first equality in (68) is easy to get using the change of variables $\nu'_k = [1 - W'(1+\delta)/W_0]^{-1/2}\nu_k$, whereas the second one is due to the fact that the supremum of the sum over $\Theta_\infty^M(\alpha_0, L_0) \setminus \{\mathbf{0}\}$ is equal to the supremum over $\Theta_\infty^M(\alpha_0, L_0)$.

Finally, for ε small enough, due to the definiton of $\psi_\varepsilon(\cdot,\cdot)$ and (62) we have $\lambda^\varepsilon \geqslant C\exp(\rho(W_0 - W'))$, and thus

$$\tau\lambda^\varepsilon \geqslant C\exp\left(\frac{\delta\rho W'}{2}\right) \longrightarrow \infty \qquad \text{as} \quad \varepsilon \to 0. \tag{69}$$

To finish the proof of the lemma it remains to substitute (68) into (67) and to take the limits of the resulting inequality first as $M \to \infty$ (using Lemma 4), then as $\varepsilon \to 0$ (using (66) and (69)), and finally as $\delta \to 0$ and $\delta_0 \to 0$.

P r o o f o f T h e o r e m 3. The assumption of the theorem guarantees that there exists $\delta \in (0,1)$ such that

$$\sup_{\theta\in\Theta(\alpha_0,L_0)} \frac{R^\varepsilon(\hat\theta,\theta)}{\psi_\varepsilon(\alpha_0,L_0)} \leqslant 1-\delta$$

for all ε small enough. Substituting this into (61) and choosing in (61) the value $\alpha' = 4\gamma_0/\delta - \rho > \alpha_0$ we get for $\alpha > \alpha'$ and for sufficiently small ε

$$\inf_{\hat\theta} \max\left\{1-\delta,\ \sup_{\theta\in\Theta(\alpha,L)} \frac{R^\varepsilon(\hat\theta,\theta)}{\psi_\varepsilon(\alpha',L)}\right\} \geqslant 1 - \frac{\delta}{4} + o(1) \geqslant 1 - \frac{\delta}{2}.$$

Thus, for ε small enough,

$$\sup_{\theta\in\Theta(\alpha,L)} \frac{R^\varepsilon(\hat\theta,\theta)}{\psi_\varepsilon(\alpha,L)} \geqslant \left(1-\frac{\delta}{2}\right)\frac{\psi_\varepsilon(\alpha',L)}{\psi_\varepsilon(\alpha,L)}. \tag{70}$$

On the other hand, it follows from [9] that the minimax risk for $\Theta(\alpha_0, L_0)$ satisfies

$$\inf_{\hat\theta}\ \sup_{\theta\in\Theta(\alpha_0,L_0)} R^\varepsilon(\hat\theta,\theta) \geqslant Cr_\varepsilon(\alpha_0,L_0), \tag{71}$$

and, in view of (20) and (21), for $r_k = q_k \equiv 1$ we have

$$\frac{r_\varepsilon(\alpha_0,L_0)}{\psi_\varepsilon(\alpha_0,L_0)} \geqslant C\left(\log\frac{1}{\varepsilon}\right)^{-\alpha_0/\gamma_0}.$$

Using the last inequality and (70), (71), we get

$$\begin{aligned}
\sup_{\theta\in\Theta(\alpha_0,L_0)} \frac{R^\varepsilon(\hat\theta,\theta)}{\psi_\varepsilon(\alpha_0,L_0)} \sup_{\theta\in\Theta(\alpha,L)} \frac{R^\varepsilon(\hat\theta,\theta)}{\psi_\varepsilon(\alpha,L)} &\geqslant C\left(\log\frac{1}{\varepsilon}\right)^{-\alpha_0/\gamma_0} \frac{\psi_\varepsilon(\alpha',L)}{\psi_\varepsilon(\alpha,L)} \\
&\geqslant C\left(\log\frac{1}{\varepsilon}\right)^{-\alpha_0/\gamma_0} \frac{W^*(\alpha',L)}{W^*(\alpha,L)} \exp\Big(\rho[W^*(\alpha',L) - W^*(\alpha,L)]\Big) \\
&\geqslant C\left(\log\frac{1}{\varepsilon}\right)^{-\alpha_0/\gamma_0+\rho(1/\gamma'-1/\gamma)} \varepsilon^{-\rho(1/\gamma'-1/\gamma)} \longrightarrow \infty
\end{aligned}$$

as $\varepsilon \to 0$.

REFERENCES

1. *Бакушинский А. Б.* О построении регулянизующих алгоритмов при случайных помехах. — Докл. АН СССР, 1969, т. 189, № 2, с. 231–233.
2. *Cavalier L.*, *Golubev G. K.*, *Picard D.*, *Tsybakov A. B.* Oracle inequalities for inverse problems. — Ann. Statist., 2002, v. 30, № 3, p. 843–874.
3. *Cavalier L.*, *Tsybakov A. B.* Sharp adaptation for inverse problems with random noise. — Probab. Theory Relat. Fields, 2002, v. 123, № 3, p. 323–354
4. *Efromovich S.* Robust and efficient recovery of a signal passed through a filter and then contaminated by nonGaussian noise. — IEEE Trans. Inf. Theory, 1997, v. 43, № 4, p. 1184–1191.
5. *Efromovich S.*, *Koltchinskii V.* On inverse problems with unknown operators. — IEEE Trans. Inform. Theory, 2001, v. 47, № 7, p. 2876–2893.
6. *Ermakov M. S.* Minimax estimation of the solution of an ill-posed convolution type problem. — Проблемы передачи информации, 2001, т. 25, № 3, p. 28–39.
7. *Goldenshluger A.*, *Pereverzev S. V.* Adaptive estimation of linear functionals in Hilbert scales from indirect white noise observations. — Probab. Theory Relat. Fields, 2000, v. 118, № 2, p. 169–186.
8. *Голубев Г. К.*, *Хасьминуский Р. З.* Статистический подход к некоторым обратным задачам для уравнений в частных производных. — Проблемы передачи информации, 1999, т. 35, в. 23, с. 51–66.
9. *Golubev G. K.*, *Khasminskii R. Z.* A statistical approach to the Cauchy problem for the Laplace equation. — State of the Art in Probability and Statistics. Festschrift for Willem R. van Zwet. Ed. by M. de Gunst, C. Klaassen, A. van der Vaart, Beachwood, OH: IMS, Institute of Mathematical Statistics, 2000, p. 419–433. (IMS Lecture notes — Monograph Series, v. 36.)
10. *Hall P.*, *Kerkyacharian G.*, *Picard D.* Block threshold rules for curve estimation using kernel and wavelet methods. — Ann. Statist., 1998, v. 26, № 3, p. 922–942.
11. *Johnstone I. M.* Wavelet shrinkage for correlated data and inverse problems: Adaptivity results. — Statistica Sinica, 1999, v. 9, № 1, p. 51–83.
12. *Лепский О. В.* Об одной задаче адаптивного оценивания в гауссовском белом шуме. — Теория вероятн. и ее примен., 1990, т. 35, в. 3, с. 459–470.
13. *Mair B.*, *Ruymgaart F. H.* Statistical inverse estimation in Hilbert scales. — SIAM J. Appl. Math., 1996, v. 56, № 5, p. 1424–1444.
14. *Mathé P.*, *Pereverzev S. V.* Optimal discretization and degrees of ill-posedness for inverse estimation in Hilbert scales in the presence of random noise. Preprint 469, WIAS, Berlin.
15. *Natterer F.* Error bounds for Tikhonov regularization in Hilbert scales. — Appl. Anal., 1984, v. 18, p. 29–37.
16. *Pensky M.*, *Vidakovic B.* Adaptive wavelet estimator for nonparametric density deconvolution. — Ann. Statist., 1999, v. 27, № 6, p. 2033–2053.
17. *Судаков В. Н.*, *Халфин Л. А.* Статистический подход к корректности задач математической физики. — Докл. АН СССР, 1964, т. 157, № 5, с. 1058–1060.
18. *Tsybakov A. B.* Pointwise and sup-norm sharp adaptive estimation of functions on the Sobolev classes. — Ann. Statist., 1998, v. 26, p. 2420–2469.
19. *Tsybakov A. B.* On the best rate of adaptive estimation in some inverse problems. — C. R. Acad. Sci. Paris, v. 330, Série I, 2000, p. 835–840.

Поступила в редакцию
23.VII.2002