Journal of Statistical Planning and Inference 139 (2009) 1821 – 1835

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

On the stability of the risk hull method for projection estimators

C. Marteau*

Institut de Mathématique, Université de Toulouse, INSA - Département GMM, 135 avenue de Rangueil, 31 077 Toulouse Cedex 4, France

ARTICLE INFO

Article history:
Received 28 March 2007
Received in revised form
24 June 2008
Accepted 3 September 2008
Available online 5 October 2008

Keywords:
Oracle inequality
Projection estimation
Risk hull
Singular value decomposition

ABSTRACT

We consider in this paper the regularization by projection of a linear inverse problem $Y = Af + \varepsilon\xi$ where ξ denotes a Gaussian white noise, A a compact operator and $\varepsilon > 0$ a noise level. Compared to the standard unbiased risk estimation (URE) method, the risk hull minimization (RHM) procedure presents a very interesting numerical behavior. However, the regularization in the singular value decomposition setting requires the knowledge of the eigenvalues of A. Here, we deal with noisy eigenvalues: only observations on this sequence are available. We study the efficiency of the RHM method in this situation. More generally, we shed light on some properties usually related to the regularization with a noisy operator.

Contents

1. Introduction

Let H and K two separable Hilbert spaces and A a linear operator acting from H to K. Our aim is to recover a function $f \in H$ from noisy measurement:

$$Y = Af + \varepsilon\xi \tag{1.1}$$

for some noise ξ and $0 < \varepsilon < 1$. The operator A is supposed to be compact. The problem is ill-posed: the solution of (1.1) does not continuously depend on the data. Only approximations, obtained via regularization methods are available. Here and in the sequel, statistical inverse problems are considered: ξ is supposed to be a Gaussian white noise (see Hida (1980) for more details).

* Tel.: +33 561 55 93 04.
E-mail address: clement.marteau@insa-toulouse.fr

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2008.09.010

The regularization of ill-posed inverse problems has been studied in both numerical and statistical areas. For a survey, we refer for instance to Engl et al. (1996) or Korostelev and Tsybakov (1993). We may also mention Johnstone and Silverman (1990) concerning tomography, Fan (1991) and Johnstone et al. (2004) for deconvolution problems or Golubev (2004) concerning the Dirichlet problem for the Laplace equation.

A regularization method is characterized by a regularization operator and a parameter choice rule. This rule may be associated to some a priori information on the solution as regularity, ℓ^2-norm, existence of jump, and so on. However, data-driven choice rules are preferable since these informations are in most cases unknown.

The well-known unbiased risk estimation (URE) procedure has been widely studied (see Section 2 for some references on the subject). Sharp results have been obtained concerning the performances of this algorithm. However, the stochastic processes involved in this method are not always well controlled. This phenomenon leads to a very unstable procedure, in particular when dealing with ill-posed inverse problems.

In this context, Cavalier and Golubev (2006) proposes a new data-driven parameter choice rule for projection (also called spectral cut-off) estimators: the risk hull minimization (RHM) method. The principle is to construct a criteria taking into account the variability of the problem. The risk hull method presents a very interesting theoretical and numerical behavior. It can also be easily implemented using Monte-Carlo approximations.

Any regularization method requires some informations on the operator A. Cavalier and Golubev (2006) assumed a complete knowledge of the spectral decomposition of A. However, in many problems and areas, the operator may only be numerically approximated or independently observed. In this case, these approximations could be used instead of the "true" operator. Nevertheless, it is not clear that such an approach may lead to a similar efficiency. In the last decade, many authors were interested in this question. We mention for instance Efromovich and Koltchinskii (2001), Cavalier and Hengartner (2005), Hoffmann and Reiß (2008) or Cavalier and Raimondo (2005). The underneath aim is to provide an accurate understanding of linear inverse problems. The following question naturally arises: 'does the RHM method is still efficient with noise in the operator?' In this paper, we give a positive answer. Moreover, we present some usual properties related to regularization with noise in the operator.

This paper is organized as follows. In Section 2, we recall the main properties of the RHM method and we propose an estimator using noisy measurements of the operator. Section 3 contains the main theorem, some numerical simulations and a brief discussion on the results. Proofs and technical lemmata are gathered in Sections 4 and 5.

2. The risk hull method

2.1. Statement of the problem

The function f has to be recovered from the noisy observation (1.1). The singular value decomposition (SVD) is a classical tool. Let $A^\star$ be the adjoint of A. The operator $A^\star A$ is compact and self-adjoint with strictly positive eigenvalues $(b_k^2)_{k\geqslant 1}$. The set of eigenfunctions $(\phi_k)_{k\geqslant 1}$ is assumed to be orthonormal. For all $k\in\mathbb{N}$, set $\psi_k = b_k^{-1}A\phi_k$. Model (1.1) can then be expressed in the sequence space form:

$$\langle Y,\psi_k\rangle = y_k = b_k\theta_k + \varepsilon\xi_k,\quad k\in\mathbb{N}, \tag{2.1}$$

where $\theta_k = \langle f,\phi_k\rangle$ for all $k\in\mathbb{N}$. The ξ_k are i.i.d. standard Gaussian random variables, since the set $(\psi_k)_{k\geqslant 1}$ is also orthonormal. In the L^2 sense, recovering the function f is equivalent to recover the sequence $\theta=(\theta_k)_{k\geqslant 1}$.

Since the operator A is compact, the sequence of eigenvalues $(b_k^2)_{k\geqslant 1}$ converges to 0 as $k\to+\infty$. The problem is ill-posed: the solution does not continuously depend on the data Y. For large values of k, the term $b_k\theta_k$ is attenuated compared to the noise $\varepsilon\xi_k$. This is the main difficulty of regularization. In this context, the projection estimation is a very simple and natural method. Let $N\in\mathbb{N}$ be fixed. The associated projection estimator (also called spectral cut-off) is defined by

$$\hat\theta^N = \sum_{k=1}^N b_k^{-1} y_k\phi_k.$$

It is equivalent to

$$\hat\theta_k^N = \begin{cases} b_k^{-1}y_k & \text{if } k\leqslant N,\\ 0 & \text{else,}\end{cases}$$

where $\hat\theta_k^N = \langle\hat\theta^N,\phi_k\rangle$. The choice of the bandwidth N is a trade-off between a good accuracy and a control of the noise in the data. It is related to the quadratic risk:

$$R(\theta,N) = \mathbb{E}_\theta\|\hat\theta^N-\theta\|^2 = \sum_{k>N}\theta_k^2 + \varepsilon^2\sum_{k=1}^N b_k^{-2}. \tag{2.2}$$

Our purpose is to construct a data-driven bandwidth $N^\star$, not depending on some a priori informations on f. We compare the risk of the associated estimator $\theta^\star$ to

$$R(\theta,N_0) = \inf_N \mathbb{E}_\theta\|\hat\theta^N-\theta\|^2.$$

The bandwidth N_0 is called the oracle. We want to construct the best possible adaptive bandwidth $N^\star$ in the sense that $R(\theta, N^\star)$ should be as close as possible to $R(\theta, N_0)$. In the particular case where $R(\theta, N^\star)$ converges to the oracle risk as $\varepsilon \to 0$, the bandwidth $N^\star$ asymptotically corresponds to the best possible choice.

2.2. Data-driven choices for the bandwidth

Several data-driven procedures have been proposed for solving this problem. Some of them are based on the well-known URE method. The principle is as follows: given a sequence θ, we would like to select the best possible bandwidth, i.e. the oracle N_0. Such a choice can only be approximated since $R(\theta, N)$ depends on θ. Hence, define

$$\tilde{N} = arg \min_{N \in \mathbb{N}} U(N, y) \quad \text{with } U(N, y) = -\sum_{l=1}^{N} b_k^{-2} y_k^2 + 2\varepsilon^2 \sum_{k=1}^{N} b_k^{-2}. \tag{2.3}$$

The quantity $U(N, y)$ is a natural estimator for the quadratic risk. This method was studied in Cavalier et al. (2002), and sharp oracle inequalities have been obtained. However, the numerical simulations are not always satisfying. Indeed, $b_k^{-1} y_k \sim \mathcal{N}(\theta_k, \varepsilon^2 b_k^{-2})$ for all $k \in \mathbb{N}$. Since we deal with ill-posed inverse problems, the variance of the unbiased risk estimator is very large. This explains in part the difficulties of the URE method. The bandwidth selection (2.3) allows large choices for the parameter N when the oracle N_0 is typically small. We refer to Cavalier and Golubev (2006) for a complete discussion and some numerical simulations.

In this context, Cavalier and Golubev (2006) introduced a new data-driven bandwidth choice rule called the RHM method. The principle is rather intuitive. Instead of the quadratic risk $R(\theta, N)$, consider the loss:

$$l(\theta, N) = \|\hat{\theta}^N - \theta\|^2 = \sum_{k=N+1}^{+\infty} \theta_k^2 + \varepsilon^2 \sum_{k=1}^{N} b_k^{-2} \xi_k^2. \tag{2.4}$$

Due to the ill-posedness of the problem, the variance of the last term in the right-hand side of (2.4) is very large. The URE procedure is based on the average behavior of $l(\theta, N)$: the variability of the problem is neglected. This explains also the instability of the URE method.

Since the sequence $(\xi_k)_{k \in \mathbb{N}}$ is unknown, construct a hull for the loss, i.e. a deterministic quantity $V(\theta, N)$ fulfilling, for all θ:

$$\mathbb{E}_\theta \sup_{N \in \mathbb{N}} [l(\theta, N) - V(\theta, N)] \leqslant 0.$$

Then estimate and minimize $V(\theta, N)$ instead of $R(\theta, N)$. This approach has been studied both from theoretical and numerical point of view. This procedure clearly improves the URE performances.

2.3. Inverse problem with noisy operator

The eigenvalues of the operator A are explicitly involved in the construction of the hull of Cavalier and Golubev (2006). This is the case for all the regularization methods constructed in the SVD framework. In many situations, A is not well known or only up to a parameter. Consider for instance a convolution operator with kernel K. It may be interesting to invert the problem using only numerical approximations or independent observations on K. This problematic is closely related to the framework of blind deconvolution, which has given rise to a considerable amount of literature. We may mention for instance Bertero et al. (1998) or Pruessner and O'Leary (2003).

In this paper, for all $k \in \mathbb{N}$, we deal with the data:

$$\begin{cases} y_k = b_k \theta_k + \varepsilon \xi_k, \\ x_k = b_k + \sigma \zeta_k, \end{cases} \tag{2.5}$$

where the ζ_k are i.i.d. standard Gaussian random variables independent of $(\xi_k)_{k \geqslant 1}$ and $\sigma \geqslant 0$ is a noise level.

The regularization of inverse problems with a noisy operator has already been considered. In the model selection context, Cavalier and Hengartner (2005) obtained a sharp oracle inequality with an estimator based on the URE method. In a different setting, we mention also the work of Efromovich and Koltchinskii (2001), Cavalier and Raimondo (2005) or Marteau (2006).

The joint asymptotic of ε, σ is very important. Here and in the sequel, we assume that $\sigma = O(\varepsilon)$ as $\varepsilon \to 0$. Hoffmann and Reiß (2008) also considered the case $\sigma \gg \varepsilon$. They proved that the rate of convergence in such a situation only depends on σ. This could be certainly generalized in our framework, but with longer proofs.

Using directly $(x_k)_{k \geqslant 1}$ instead of $(b_k)_{k \geqslant 1}$ is a natural but naive idea. Due to the ill-posedness of the problem, the sequence $(b_k)_{k \geqslant 1}$ converges to 0 as $k \to +\infty$. For large values of k, there is mainly noise in x_k: this is a problem when estimating b_k^{-1}. Hence, define the following stopping rule:

$$M = \min \left\{ k \leqslant T_0 : |x_k| \leqslant \sigma \log^{3/2} \frac{1}{\sigma} \right\} - 1, \tag{2.6}$$

where T_0 ensures that M is not too large. For example, choose $T_0=\sigma^{-2}$. The quantity M is stochastic but can be controlled. Cavalier and Hengartner (2005) proved that

$$M_0 < M < M_1,$$

with a large probability, where

$$M_0=\min\left\{k : |b_k|\leqslant \sigma \log^2\frac{1}{\sigma}\right\}, \tag{2.7}$$

and

$$M_1=\min\left\{k : |b_k|\leqslant \sigma \log\frac{1}{\sigma}\right\}. \tag{2.8}$$

From Lemma 5 in Section 4, x_k is close to b_k with a large probability for $k\leqslant M$. For $k>M$, there is too much noise in x_k in order to provide a good quality of estimation. Hence, we will only estimate the first M coefficients of the sequence $(\theta_k)_{k\in\mathbb{N}}$. This can be a problem in some particular cases: see Section 3 for a complete discussion.

We are now ready to construct an estimator of θ. For all $N\in\mathbb{N}$, define

$$U_0(N)=\inf\{t>0 : \mathbb{E}\eta_N\mathbf{1}_{\{\eta_N\geqslant t\}}\leqslant\varepsilon^2\}\quad \text{with } \eta_N=\varepsilon^2\sum_{k=1}^{N}b_k^{-2}(\xi_k^2-1).$$

Theorem 2 of Cavalier and Golubev (2006) provides that the quantity

$$V_\alpha(\theta,N)=\sum_{k=N+1}^{+\infty}\theta_k^2+\varepsilon^2\sum_{k=1}^{N}b_k^{-2}+(1+\alpha)U_0(N)+\frac{C^\star\varepsilon^2}{\alpha} \tag{2.9}$$

is a risk hull for all $\alpha>0$, i.e.

$$\mathbb{E}_\theta\sup_N[l(\theta,N)-V_\alpha(\theta,N)]\leqslant 0,$$

where $l(\theta,N)$ is defined in (2.4) and $C^\star$ denotes a positive constant independent of ε. Cavalier and Golubev (2006) proposes to minimize the following estimator of $V_\alpha(\theta,N)$:

$$V_\alpha(y,N)=-\sum_{k=1}^{N}b_k^{-2}y_k^2+2\varepsilon^2\sum_{k=1}^{N}b_k^{-2}+(1+\alpha)U_0(N). \tag{2.10}$$

Since the sequence $(b_k)_{k\geqslant 1}$ is unknown here, consider instead

$$V_\alpha(x,y,N)=-\sum_{k=1}^{N}x_k^{-2}y_k^2+2\varepsilon^2\sum_{k=1}^{N}x_k^{-2}+(1+\alpha)U_0(N,x), \tag{2.11}$$

where

$$U_0(N,x)=\inf\{t>0 : \mathbb{E}_x\tilde\eta_N\mathbf{1}_{\{\tilde\eta_N\geqslant t\}}\leqslant\varepsilon^2\},\ \ \tilde\eta_N=\varepsilon^2\sum_{k=1}^{N}x_k^{-2}(\xi_k^2-1),$$

and $\mathbb{E}_x$ denotes the conditional expectation on the sequence $(x_k)_{k\geqslant 1}$. We prove in Section 3 that the estimator:

$$\theta^\star=\sum_{k=1}^{N^\star}x_k^{-1}y_k\phi_k, \tag{2.12}$$

associated to the bandwidth

$$N^\star=arg\min_{N\leqslant M}V_\alpha(x,y,N) \tag{2.13}$$

is a relevant choice from theoretical and numerical point of view.

The risk hull constructed in Cavalier and Golubev (2006) is in fact a penalized URE method. The penalty is chosen in order to contain the stochastic terms in (2.3) and (2.4). This penalty is directly computable via Monte-Carlo approximations.

In the inverse problems framework, different penalizations have been proposed by, for instance, Golubev (2004), Loubes et al. (2005) or Loubes and Ludena (2007). In the direct case (i.e. when A denotes the identity operator), we may also mention Barron et al. (1999) or Birgé and Massart (2001).

3. Theoretical results and simulations

3.1. Main result

Let v a given real sequence and $r \in \mathbb{R}$. Here and in the sequel, we write $v \sim (k^r)_{k\geqslant 1}$ if we can find d_0 and d_1 positive constants fulfilling: $d_0 k^r \leqslant v_k \leqslant d_1 k^r$ for all $k \in \mathbb{N}$.

From now, we assume a polynomial decay for the eigenvalues of the operator: $(b_k)_{k\geqslant 1} \sim (k^{-\beta})_{k\geqslant 1}$ for some $\beta \geqslant 0$. The problem is said to be mildly ill-posed.

For all $N \in \mathbb{N}$, define

$$R_\alpha(\theta, N) = \sum_{k=N+1}^{+\infty} \theta_k^2 + \varepsilon^2 \sum_{k=1}^{N} b_k^{-2} + (1+\alpha) U_0(N). \tag{3.1}$$

The two associated oracle bandwidths are

$$\bar{N}_0 = arg \inf_N R_\alpha(\theta, N) \quad \text{and} \quad \tilde{N}_0 = arg \inf_{N \leqslant M} R_\alpha(\theta, N). \tag{3.2}$$

The proof of the following theorem is given in Section 5.

Theorem 1. *Let $\theta^\star$ the estimator defined in* (2.12) *and* (2.13). *Assume that $(b_k)_{k\in\mathbb{N}} \sim (k^{-\beta})_{k\in\mathbb{N}}$ for some $\beta > 0$. There exist positive constants γ_0, C_0, C_1 and τ such that $\forall \gamma \in (0, \gamma_0]$ and $\alpha > 1$:*

$$\mathbb{E}_\theta \|\theta^\star - \theta\|^2 \leqslant (1 + \varphi_\sigma(\gamma)) \inf_N R_\alpha(\theta, N) + C_1 \frac{\varepsilon^2}{\gamma^{4\beta+1}} \left(1 + \frac{\sigma^2}{\varepsilon^2}\|\theta\|^2\right)^{2\beta+1} + \frac{C_1 \varepsilon^2}{(\alpha - 1)} + \Gamma(\theta) + \Omega, \tag{3.3}$$

where $\Omega = C_0 M_1 (1 + \|\theta\|^2) \exp(-\log^{1+\tau} 1/\sigma)$, $\varphi_\sigma(\gamma)$ vanishes as $\gamma, \sigma \to 0$ and

$$\Gamma(\theta) = \sum_{k=M_0}^{M_0 \vee \bar{N}_0} \theta_k^2 + \varepsilon^2 \sum_{k=M_0}^{M_0 \vee \bar{N}_0} b_k^{-2} + (1+\alpha)[U_0(\bar{N}_0 \vee M_0) - U_0(M_0)]. \tag{3.4}$$

By convention $\sum_a^a = 0$ and $a \vee b = \max(a, b)$ for all $a, b > 0$.

Our results always hold with the polynomial hypotheses of Cavalier and Golubev (2006). However, for the sake of convenience, we prefer to use simpler assumptions concerning the sequence $(b_k)_{k\in\mathbb{N}}$.

The risk of our estimator is close to the best possible one (the oracle). This result differs from Theorem 1 of Cavalier and Golubev (2006) by some residual terms depending of σ and on the norm of f. For large values of $\|\theta\|$, the quality of estimation will then be rather bad. This fact is not a specificity of the risk hull method. Such a residual term appears in all the situations where noisy measurements of the operator are involved. We refer for instance to the proofs of Cavalier and Hengartner (2005), Efromovich and Koltchinskii (2001), Hoffmann and Reiß (2008) or Marteau (2006). This phenomenon is in fact rather simple. When $\sigma = 0$, we use the data

$$b_k^{-1} y_k = \theta_k + \varepsilon b_k^{-1} \xi_k, \tag{3.5}$$

in order to estimate each coefficient θ_k. With noise in the operator, we use instead:

$$x_k^{-1} y_k = x_k^{-1}(b_k \theta_k + \varepsilon \xi_k) = a_k \theta_k + \varepsilon x_k^{-1} \xi_k = \theta_k + \varepsilon x_k^{-1} \xi_k + (a_k - 1)\theta_k,$$

where $a_k = x_k^{-1} b_k$. With a large probability, Lemma 5 provides that $(a_k - 1)$ is close to $\sigma b_k^{-1} \zeta_k$ and x_k is of order b_k for $k \leqslant M$. Therefore, from a heuristic point of view:

$$x_k^{-1} y_k = \theta_k + \varepsilon b_k^{-1} \xi_k + \sigma b_k^{-1} \zeta_k \theta_k, \quad \forall k \leqslant M. \tag{3.6}$$

The variance of $x_k^{-1} y_k$ is then of order $\varepsilon^2 b_k^{-2} + \sigma^2 \theta_k^2 b_k^{-2}$. It increases with the norm of θ. Noise in the operator leads to a new perturbation in the data with variance depending on $\|\theta\|$. The associated error term is not always negligible. This will be illustrated via numerical simulations.

The terms Ω and $\Gamma(\theta)$ are usually associated to the regularization with a noisy operator. The term $\Gamma(\theta)$ corresponds to the residual of $R_\alpha(\theta, \bar{N}_0)$ truncated at order M where $\bar{N}_0$ is the oracle defined in (3.2). It vanishes if $M_0 > \bar{N}_0$. The behavior of this term has been studied in Cavalier and Hengartner (2005). Under some reasonable constraints, it is negligible from a minimax point of view.

The quantity Ω is exponentially decreasing as $\sigma \to 0$. We can expect that this term will not have a great influence on the performances of $\theta^\star$.

3.2. Numerical simulations

This section illustrates inequality (3.3) via some numerical simulations. In this paper, we are interested in both the RHM stability and the regularization with a noisy operator. We focus on the restrictions resulting from the presence of noise in the operator. In particular, we are interested in the link between the performances of $\theta^\star$ and the respective values of $\|\theta\|$, σ and ε. These restrictions are certainly the same, whatever the parameter choice rule. Our aim is not to provide a comparison between the RHM method and some alternative adaptive procedures, but rather to shed light on the consequences of (3.6).

In the sequel, we consider the family of functions:

$$\theta_k^a = \frac{a}{1+(k/W)^m}, \quad k \in \mathbb{N}, \tag{3.7}$$

where a is called amplitude, W bandwidth and m smoothness. Clearly, it is impossible to provide a complete numerical study and we have to restrict ourselves to some particular families of function. Set (3.7) can easily be handled. The parameters a and m have a direct influence on the quantities of interest.

From now, fix $W=m=6$. The results for different values are essentially the same and will not be reproduced here. The quality of an estimator $\bar{\theta}$ is measured by its oracle efficiency:

$$r_\varepsilon(\theta, \bar{\theta}) = \frac{\inf_N \mathbb{E}_\theta \|\hat{\theta}^N - \theta\|^2}{\mathbb{E}_\theta\|\bar{\theta}-\theta\|^2} = \frac{\mathbb{E}_\theta \|\hat{\theta}^{N_0} - \theta\|^2}{\mathbb{E}_\theta\|\bar{\theta}-\theta\|^2}. \tag{3.8}$$

Let $\tilde{\theta}$ the estimator of Cavalier and Golubev (2006) and $\tilde{N}$ the associated bandwidth. We are interested in the comparison between $r_\varepsilon(\theta, \theta^\star)$ and $r_\varepsilon(\theta, \tilde{\theta})$. We also compare the mean bandwidth $\mathbb{E}_\theta[N^\star]$ and $\mathbb{E}_\theta[\tilde{N}]$ to the oracle N_0. These quantities are computed via the Monte-Carlo method with 10 000 replications.

The relation between the oracle efficiency of $\theta^\star$ and $(\sigma^2/\varepsilon^2)\|\theta\|^2$ is also rather interesting. This term is associated to the noise in the operator. In (2.12), we use x_k^{-1} instead of b_k^{-1} in order to estimate each coefficient θ_k. The associated residual error is in fact linked to

$$d_\varepsilon(x,\theta) = \mathbb{E}_\theta\|\theta^\star - \theta\|^2 - \mathbb{E}_\theta\|\hat{\theta}^\star - \theta\|^2, \tag{3.9}$$

where

$$\hat{\theta}^\star = \sum_{k=1}^{N^\star} b_k^{-1} y_k \phi_k,$$

and $N^\star$ is defined in (2.13). This residual error will be compared to the oracle risk.

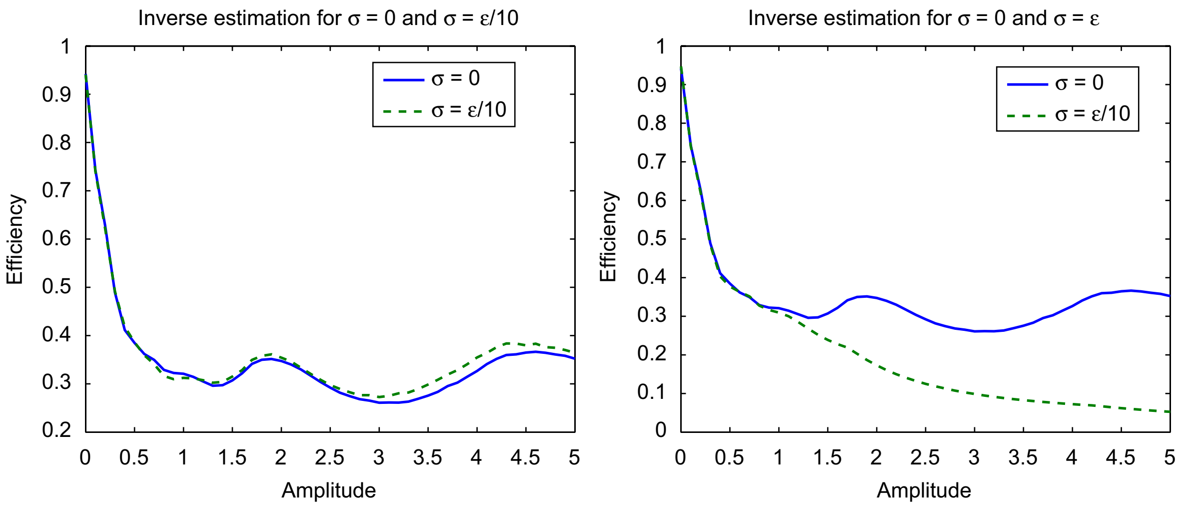

Fig. 1. Efficiency of the RHM method for $\beta = 2$.

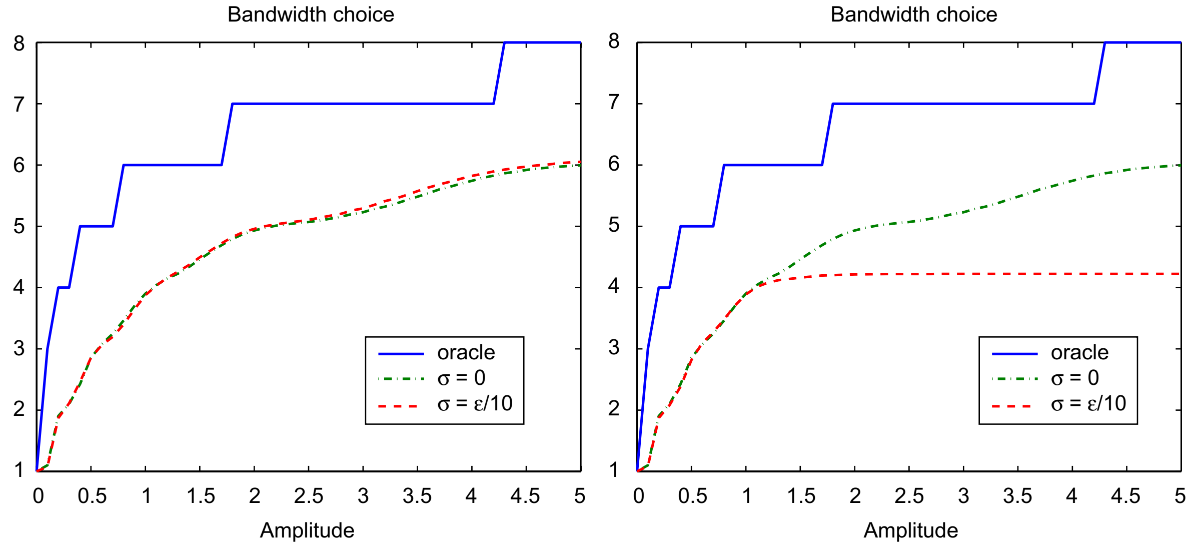

Fig. 2. Mean bandwidths.

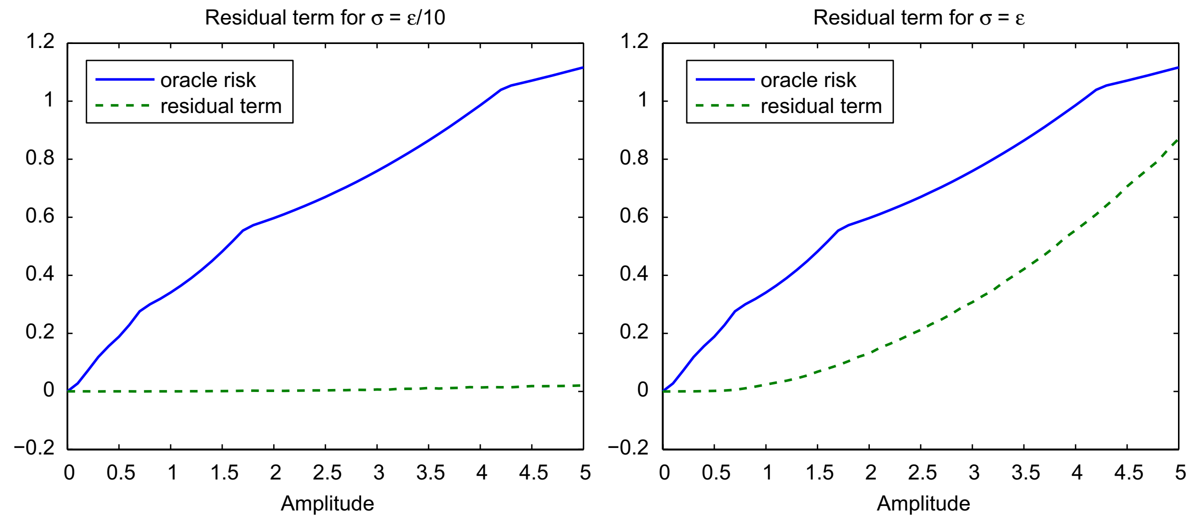

Fig. 3. Behavior of the residual term.

From now, fix $\varepsilon = 0.01$ and consider two different levels of noise in the eigenvalues: $\sigma = \varepsilon/10$ and ε. Each case is studied with $\beta = 2$ ($b_k = k^{-2}$ for all $k \in \mathbb{N}$). Results are presented in Figs. 1–3, on the l.h.s. for the case $\sigma = \varepsilon/10$ and on the r.h.s. for the case $\sigma = \varepsilon$.

When the noise in the operator is significantly small ($\sigma = \varepsilon/10$), the numerical behavior of the RHM method is rather close to the case $\sigma = 0$ for $a \in [0; 5]$. The bandwidth choice is unaffected. The term $d_\varepsilon(x, \theta)$ has no influence on the quality of estimation. It is clearly negligible compared to the oracle risk.

Results for the case $\sigma = \varepsilon$ are rather different. If a is too large (or equivalently $\|\theta\|$), the quality of estimation is really affected by the noise σ. Two different explanations are available. First, replacing b_k^{-1} by x_k^{-1} affects the risk when k is large. The quantity $d_\varepsilon(x, \theta)$, associated to the residual term $(\sigma^2/\varepsilon^2)\|\theta\|^2$ in (3.3), is not negligible for large values of a. It is of order of the optimal risk (r.h.s. of Fig. 3).

The second problem is related to the stopping rule M. It is smaller than the oracle bandwidth N_0 for large values of $\|\theta\|$ (r.h.s. of Fig. 2). Since $N^\star \leqslant M$, the number of estimated coefficients of θ is not sufficient. The corresponding error is associated to $\Gamma(\theta)$ in Theorem 1. This explain in part why we do not consider estimation for larger value of σ than 0.01. Never more than 4 or 5 coefficients are estimated. In this case, all the procedures will produce more or less similar results.

3.3. Conclusion

The control of the norm of θ is crucial when dealing with a noisy operator. This can be explained in part by (3.6). The noise in the operator leads to a new perturbation in the observations. It would be interesting to develop methods taking into account this particularity.

The term $\Gamma(\theta)$ is also rather important. It is related to the respective values of M_0 and N_0. If N_0 is too large compared to M_0, we cannot estimate a sufficient number of coefficients θ_k. Fortunately, in most cases, $N_0/M_0 \to 0$ as $\varepsilon, \sigma \to 0$. In some sense, recovering the eigenvalues $(b_k)_{k\in\mathbb{N}}$ is easier than the sequence θ. The regularization with noise in the operator is clearly asymptotic (in an oracle point of view). This is rather different than for the case $\sigma = 0$. All these properties are not specific to the RHM method but rather to the regularization with noise in the operator.

We conclude this paper with a short discussion on the parameter α. In Cavalier and Golubev (2006), it seems that the condition $\alpha > 1$ is too restrictive. Indeed, acceptable bounds of the risk are available for $\sigma = 0$ and $0 < \alpha \leqslant 1$. For $\sigma > 0$, the setting is a little bit different. In this case, the condition $\alpha > 1$ enables us to control the difference between $U_0(N)$ and $U_0(N,x)$ (see Lemma 3 and inequality (5.13)). The noisy penalty is large enough to contain the stochastic term in $V_\alpha(x,y,N)$. The RHM procedure is stable.

4. Technical lemmas

From now, we consider $\Omega = O(M_1 \exp[-\log^{1+\tau} 1/\sigma])$ as $\sigma \to 0$. Therefore, we often write $\Omega + \Omega = \Omega$. The resulting constant C_0 in Theorem 1 is the sum of all the constants appearing in the proof. For all $N \in \mathbb{N}$, introduce also the term:

$$\Delta(N) = \sup_{k=1,\dots,N} \varepsilon^2 b_k^{-2}, \tag{4.1}$$

and the event:

$$\mathscr{B} = \bigcap_{k=1}^{M} \{\sigma |b_k^{-1} \zeta_k| \leqslant \nu\sqrt{2} - 1\}, \tag{4.2}$$

where $1/\sqrt{2} < \nu < 1$. On this event, $a_k = x_k^{-1} b_k$ is close to 1 uniformly in $k \leqslant M$. Moreover, Lemma 5 provides that $P(\mathscr{B}^c)$ decreases exponentially fast as $\sigma \to 0$. Hence, x_k can easily be controlled with a large probability for $k \leqslant M$.

Lemma 1. *Let $\theta^\star$ defined in* (2.12) *and* (2.13). *There exists $\rho_\sigma \to 0$ as $\sigma \to 0$ fulfilling*:

$$\mathbb{E}_\theta \|\theta^\star - \theta\|^2 \leqslant (1+\rho_\sigma)\mathbb{E}_\theta l(\theta, N^\star) + C c_\sigma^2 \|\theta\|^2 \frac{\sigma^2}{\varepsilon^2} \mathbb{E}_\theta \Delta(N^\star) + \Omega,$$

where $C > 0$, $c_\sigma = \log^{4/3} 1/\sigma$ and $l(\theta, N^\star)$ is defined in (2.4).

Proof. For all $k \in \mathbb{N}$, set $a_k = x_k^{-1} b_k$. By simple algebra:

$$\begin{aligned}
\mathbb{E}_\theta \|\theta^\star - \theta\|^2 &= \mathbb{E}_\theta \left[\sum_{k=1}^{+\infty} (\theta_k^\star - \theta_k)^2 \right] = \mathbb{E}_\theta \left[\sum_{k=1}^{N^\star} (x_k^{-1} y_k - \theta_k)^2 \right] + \mathbb{E}_\theta \left[\sum_{k=N^\star+1}^{+\infty} \theta_k^2 \right] \\
&= \mathbb{E}_\theta \left[\sum_{k=1}^{N^\star} (a_k \theta_k + \varepsilon x_k^{-1} \xi_k - \theta_k)^2 \right] + \mathbb{E}_\theta \left[\sum_{k=N^\star+1}^{+\infty} \theta_k^2 \right] \\
&= \mathbb{E}_\theta \left[\sum_{k=1}^{N^\star} \{\theta_k^2 (1-a_k)^2 + \varepsilon^2 x_k^{-2} \xi_k^2 + 2\varepsilon (a_k - 1)\theta_k x_k^{-1} \xi_k\} \right] + \mathbb{E}_\theta \sum_{k>N^\star} \theta_k^2.
\end{aligned}$$

Using (2.4),

$$\begin{aligned}
\mathbb{E}_\theta \|\theta^\star - \theta\|^2 - \mathbb{E}_\theta l(\theta, N^\star) &= \mathbb{E}_\theta \left[\sum_{k=1}^{N^\star} \theta_k^2 (1-a_k)^2 \right] + \mathbb{E}_\theta \left[\sum_{k=1}^{N^\star} \varepsilon^2 b_k^{-2} (a_k^2 - 1)\xi_k^2 \right] + 2\varepsilon \mathbb{E}_\theta \left[\sum_{k=1}^{N^\star} (1-a_k)\theta_k x_k^{-1} \xi_k \right] \\
&= A_1 + A_2 + A_3.
\end{aligned} \tag{4.3}$$

Begin with the study of A_1:

$$(1-a_k)^2 \mathbf{1}_{\mathscr{B}} = \frac{(\sigma b_k^{-1} \zeta_k)^2}{(1+\sigma b_k^{-1} \zeta_k)^2} \mathbf{1}_{\mathscr{B}} \leqslant c\sigma^2 b_k^{-2} \zeta_k^2 \, \mathbf{1}_{\mathscr{B}}. \tag{4.4}$$

Introduce $c_\sigma = \log^{4/3} 1/\sigma$. Using (4.4), Lemma 5 and the inequality $N^\star \leqslant M$, we obtain

$$\begin{aligned} A_1 &= \mathbb{E}_\theta \sum_{k=1}^{N^\star} \theta_k^2 (1-a_k)^2 = \mathbb{E}_\theta \left[\sum_{k=1}^{N^\star} \theta_k^2 (1-a_k)^2 1_{\mathscr{B}} \right] + \mathbb{E}_\theta \left[\sum_{k=1}^{N^\star} \theta_k^2 (1-a_k)^2 1_{\mathscr{B}^c} \right] \\ &\leqslant C\|\theta\|^2 \mathbb{E}_\theta \left[\sup_{i=1,\dots,N^\star} \frac{\sigma^2}{b_i^2} \zeta_i^2 \right] + CM_1 \|\theta\|^2 \,\mathrm{e}^{-\log^{1+\tau} 1/\sigma} \leqslant c\|\theta\|^2 c_\sigma \frac{\sigma^2}{\varepsilon^2} \mathbb{E}_\theta \left[\Delta(N^\star) \right] + \Omega \end{aligned} \tag{4.5}$$

for some $\tau > 0$, slightly smaller than in Lemma 5. We have used the large deviation inequality:

$$P\left(\sup_{i=1,\dots,N^\star} \zeta_i^2 > c_\sigma \right) \leqslant M_1 P(|\zeta_1| > \sqrt{c_\sigma}) + \mathrm{e}^{-\log^{1+\tau} 1/\sigma} \leqslant C \,\mathrm{e}^{-\log^{1+\tau} 1/\sigma}. \tag{4.6}$$

By the same way, using (ii) of Lemma 5:

$$\begin{aligned} A_2 &= \left| \mathbb{E}_\theta \sum_{k=1}^{N^\star} \varepsilon^2 b_k^{-2} (a_k^2 - 1) \xi_k^2 \right| \leqslant \mathbb{E}_\theta \left[\sum_{k=1}^{N^\star} \varepsilon^2 b_k^{-2} \xi_k^2 \left| \left(-2 \frac{\sigma}{b_k} \zeta_k + 6 \frac{\sigma^2}{b_k^2} \zeta_k^2 \right) \right| \right] + \Omega \\ &\leqslant c \frac{\sqrt{c_\sigma}}{\log 1/\sigma} \mathbb{E}_\theta \left[\sum_{k=1}^{N^\star} \varepsilon^2 b_k^{-2} \xi_k^2 \right] + \Omega \leqslant c \left(\log^{-1/3} \frac{1}{\sigma} \right) \mathbb{E}_\theta [l(\theta, N^\star)] + \Omega. \end{aligned} \tag{4.7}$$

To finish the proof, apply the elementary inequality $2ab \leqslant Ba^2 + B^{-1} b^2$ to

$$A_3 = 2\varepsilon \left| \mathbb{E}_\theta \sum_{k=1}^{N^\star} (1-a_k) \theta_k x_k^{-1} \xi_k \right| \leqslant BA_1 + cB^{-1} \mathbb{E}_\theta \left[l(\theta, N^\star) \right] + \Omega.$$

Using (4.3)–(4.7), we eventually obtain the lemma, when choosing for instance $B = c_\sigma$. $\square$

Lemma 2. *Assume $(b_k)_{k\geqslant 1} \sim (k^{-\beta})_{k \geqslant 1}$ for some $\beta \geqslant 0$. Let $V_\mu(\theta, N^\star)$ and $\Delta(N^\star)$ the quantities defined in* (2.9) *and* (4.1). *For all $B > 0$ and $\mu > 0$,*

$$\mathbb{E}_\theta \Delta(N^\star) \leqslant B \mathbb{E}_\theta V_\mu(\theta, N^\star) + C\varepsilon^2 B^{-2\beta}$$

for some constant $C > 0$ independent of ε.

Proof. Let $B > 0$ be fixed:

$$\begin{aligned} \mathbb{E}_\theta \Delta(N^\star) &= \mathbb{E}_\theta \sup_{k=1,\dots,N^\star} \varepsilon^2 b_k^{-2}, = \mathbb{E}_\theta \sup_{k=1,\dots,N^\star} \varepsilon^2 b_k^{-2} \mathbf{1}_{\left\{ B^{-1} \sup_{k \leqslant N^\star} \varepsilon^2 b_k^{-2} < \sum_{k=1}^{N^\star} \varepsilon^2 b_k^{-2} \right\}} \\ &\quad + \mathbb{E}_\theta \sup_{k=1,\dots,N^\star} \varepsilon^2 b_k^{-2} \mathbf{1}_{\left\{ B^{-1} \sup_{k \leqslant N^\star} \varepsilon^2 b_k^{-2} \geqslant \sum_{k=1}^{N^\star} \varepsilon^2 b_k^{-2} \right\}} \\ &\leqslant B\varepsilon^2 \mathbb{E}_\theta \sum_{k=1}^{N^\star} b_k^{-2} + C\varepsilon^2 B^{-2\beta}, \leqslant B \mathbb{E}_\theta V_\mu(\theta, N^\star) + C\varepsilon^2 B^{-2\beta}. \quad \square \end{aligned}$$

Lemma 3. *There exists $1 > g_\sigma > 0$ such that,*

$$\mathbb{E} \sup_{k \leqslant M} [U_0(k) - (1+g_\sigma) U_0(k,x)] \leqslant C\varepsilon^2 + \Omega,$$

and $g_\sigma \to 0$ as $\sigma \to 0$.

Proof. Apply Lemma 1 of Cavalier and Golubev (2006) to $U_0(k,x)$, conditioning on the sequence $(x_k)_{k \geqslant 1}$. There exists $k_0 > 0$ depending only on the sequence $(b_k)_{k \in \mathbb{N}}$, such that for all $M > k > k_0$,

$$\frac{U_0(k,x)}{\sqrt{2\tilde{\Sigma}_k}} \mathbf{1}_{\mathscr{B}} \geqslant \sqrt{\log \left(\frac{\tilde{\Sigma}_k}{2\pi \varepsilon^4} \right)} \mathbf{1}_{\mathscr{B}}, \tag{4.8}$$

where $\tilde{\Sigma}_k = \varepsilon^4 \sum_{s=1}^k x_s^{-4}$. Indeed, on the event $\mathscr{B}$ defined in (4.2), x_k is of order b_k for all $k \leqslant M$. The polynomial hypotheses of Cavalier and Golubev (2006) are satisfied by the sequence $(x_k)_{k \geqslant 1}$. Lemma 1 of the related paper can directly be applied.

Define also $\Sigma_k=\varepsilon^4\sum_{s=1}^k b_s^{-4}$ and $K_k=\eta_k/\sqrt{2\Sigma_k}$. Let $k\geqslant k_0$ be fixed. We first show that on the event $\mathscr{B}$, $U_0(k)<2U_0(k,x)$. Remark that

$$U_0(k)<2U_0(k,x)\mathbf{1}_{\mathscr{B}} \Leftrightarrow \mathbf{1}_{\mathscr{B}}\mathbb{E}_x\eta_k\mathbf{1}_{\{\eta_k\geqslant 2U_0(k,x)\}}<\varepsilon^2 \Leftrightarrow \mathbf{1}_{\mathscr{B}}\mathbb{E}_xK_k\mathbf{1}_{\{K_k\geqslant 2U_0(k,x)/\sqrt{2\Sigma_k}\}}<\frac{\varepsilon^2}{\sqrt{2\Sigma_k}}.$$

Using (4.2) and (4.8),

$$\frac{2U_0(k,x)}{\sqrt{2\Sigma_k}}\mathbf{1}_{\mathscr{B}}=\frac{2U_0(k,x)}{\sqrt{2\tilde{\Sigma}_k}}\frac{\sqrt{\tilde{\Sigma}_k}}{\sqrt{\Sigma_k}}\mathbf{1}_{\mathscr{B}}\geqslant\frac{1}{v^2}\sqrt{\log\left(\frac{\tilde{\Sigma}_k}{2\pi\varepsilon^4}\right)}\mathbf{1}_{\mathscr{B}}. \tag{4.9}$$

Denotes by $u_1(k,x)$ the last term in the r.h.s. of (4.9). From the above equation, it suffices to show

$$\mathbf{1}_{\mathscr{B}}\mathbb{E}_xK_k\mathbf{1}_{\{K_k\geqslant u_1(k,x)\}}<\frac{\varepsilon^2}{\sqrt{2\Sigma_k}}. \tag{4.10}$$

Let $t>0$, possibly depending on the sequence $(x_k)_{k\geqslant 1}$. Integrating by parts:

$$\mathbf{1}_{\mathscr{B}}\mathbb{E}_xK_k\mathbf{1}_{\{K_k\geqslant t\}}=\left[tP_x(K_k\geqslant t)+\int_t^{+\infty}P_x(K_k>u)\,\mathrm{d}u\right]\mathbf{1}_{\mathscr{B}}, \tag{4.11}$$

where P_x denotes the conditional probability on the sequence $(x_k)_{k\geqslant 1}$. Each K_k can be approximated by a standard Gaussian random variable on small intervals. Indeed, for all $m>0$, conditioning on the sequence $(x_k)_{k\geqslant 1}$, there exists a measure P_k^m verifying:

$$\max_{t_0\leqslant u\leqslant t_1}|P_k^m(u)-\phi(u)|\mathbf{1}_{\mathscr{B}}\leqslant\frac{C\varepsilon^2t_1^2}{\sqrt{2k\Sigma_k}}\mathbf{1}_{\mathscr{B}},$$

where ϕ is the density of a standard Gaussian random variable and

$$|P_x(K_k>u)-P_k^m(u)|\mathbf{1}_{\mathscr{B}}\leqslant\frac{C}{k^{m/2}}\mathbf{1}_{\mathscr{B}}.$$

See Cavalier and Golubev (2006) for more details. Using (4.11), for all $t_1\geqslant t$,

$$\begin{aligned}\mathbb{E}_xK_k\mathbf{1}_{\{K_k\geqslant t\}}\mathbf{1}_{\mathscr{B}}&=\left[tP_x(K_k\geqslant t)+\int_t^{t_1}P_x(K_k>u)\,\mathrm{d}u+\int_{t_1}^{+\infty}P_x(K_k>u)\,\mathrm{d}u\right]\mathbf{1}_{\mathscr{B}}\\&\leqslant\left[tP_k^m(t)+\int_t^{t_1}P_k^m(u)\,\mathrm{d}u+\int_{t_1}^{+\infty}P_x(K_k>u)\,\mathrm{d}u+\frac{Ct_1}{k^{m/2}}\right]\mathbf{1}_{\mathscr{B}}\\&\leqslant\left[t\phi(t)+\int_t^{+\infty}\phi(u)\,\mathrm{d}u+t_1\max_{t\leqslant u\leqslant t_1}|P_k^m(u)-\phi(u)|+\int_{t_1}^{+\infty}P_x(K_k>u)\,\mathrm{d}u+\frac{Ct_1}{k^{m/2}}\right]\mathbf{1}_{\mathscr{B}}.\end{aligned} \tag{4.12}$$

Replacing t by $u_1(k,x)$, and integrating by parts,

$$\begin{aligned}\mathbb{E}_xK_k\mathbf{1}_{\{K_k\geqslant u_1(k,x)\}}\mathbf{1}_{\mathscr{B}}&\leqslant\left[\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-u_1(k,x)^2/2}+\int_{t_1}^{+\infty}P_x(K_k>u)\,\mathrm{d}u+t_1\max_{u_1(k,x)\leqslant u\leqslant t_1}|P_k^m(u)-\phi(u)|+\frac{Ct_1}{k^{m/2}}\right]\mathbf{1}_{\mathscr{B}}\\&=T_1+T_2+T_3+T_4.\end{aligned}$$

Begin with the majoration of T_1:

$$T_1=\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-u_1(k,x)^2/2}\mathbf{1}_{\mathscr{B}}\leqslant c\frac{\varepsilon^{4/2v^4}}{\Sigma_k^{1/2v^4}}=\mathrm{o}\left(\frac{\varepsilon^2}{\sqrt{2\Sigma_k}}\right)\quad\text{as } k\to+\infty. \tag{4.13}$$

Indeed $1/\sqrt{2}<v<1$. For the study of T_2, write

$$K_k=\sum_{s=1}^k d_s^2(\xi_s^2-1),$$

where $d_s^2=\sigma_s^2/\sqrt{2\Sigma_k}$ for all $s\in\{1,\dots,k\}$. A large deviation inequality provides

$$P_x(K_k>u)\leqslant\exp\left[-\frac{u^2}{4(\sum_{s=1}^k d_s^4+u\max_{s=1,\dots,k}d_s^2)}\right]=\exp\left[-\frac{u^2}{2+4u\sigma_k^2/\sqrt{2\Sigma_k})}\right]\leqslant\exp(-u/6)$$

for all $u > 1$. Choose $t_1 = (6u_1(k,x)^2 \wedge 1)$ to obtain

$$T_2 \leqslant \int_{6u_1(k,x)^2}^{+\infty} \mathrm{e}^{-u/6}\,\mathrm{d}u\,\mathbf{1}_{\mathscr{B}} \leqslant C\exp[-u_1(k,x)^2] = \mathrm{o}\left(\frac{\varepsilon^2}{\sqrt{2\Sigma_k}}\right) \quad \text{as k} \to +\infty. \tag{4.14}$$

Indeed, the behavior of the sequence $(b_k)_{k\in\mathbb{N}}$ ensures

$$\sqrt{\log\left(\frac{k}{2\pi}\right)} \leqslant u_1(k,x)\mathbf{1}_{\mathscr{B}} \leqslant C\sqrt{\log(k)}. \tag{4.15}$$

With this choice of t_1 and (4.15), we can easily obtain a bound for the remainder terms T_3 and T_4:

$$T_3 = t_1 \max_{t\leqslant u\leqslant t_1} |P_k^m(u) - \phi(u)|\mathbf{1}_{\mathscr{B}} \leqslant \frac{C\varepsilon^2 u_1(k,x)^6}{\sqrt{2k\Sigma_k}}\mathbf{1}_{\mathscr{B}} = \mathrm{o}\left(\frac{\varepsilon^2}{\sqrt{2\Sigma_k}}\right) \quad \text{as } k \to +\infty. \tag{4.16}$$

To finish, choosing large enough m,

$$T_4 = \mathrm{o}\left(\frac{\varepsilon^2}{\sqrt{2\Sigma_k}}\right) \quad \text{as k} \to +\infty. \tag{4.17}$$

Hence, we can find k_1 independent of ε and σ verifying

$$U_0(k,x)\mathbf{1}_{\mathscr{B}} > \frac{U_0(k)}{2}, \quad \forall k \geqslant k_1.$$

Now, let $k \leqslant k_1$:

$$U_0(k) \leqslant C\varepsilon^2.$$

Indeed, k_1 is independent of ε. Then remark

$$\mathbb{E}\sup_{k\leqslant M}[U_0(k) - 2U_0(k,x)]\mathbf{1}_{\mathscr{B}^c} \leqslant \sup_{k\leqslant M}[U_0(k)] \times P(\mathscr{B}^c) + \Omega \leqslant \Omega. \tag{4.18}$$

Indeed, looking at the proof of Lemma 1 of Cavalier and Golubev (2006), we can find a convenient bound for $U_0(k)$, $k \leqslant M_1$. Approximating M_1 by $\sigma^{-2/(2\beta+1)}$, it is then easy to see that (4.18) is satisfied.

Lemma 3 is verified for $g_\sigma = 1$. Nevertheless, the proof can easily be improved. Indeed, using Taylor expansion of Lemma 5, $\tilde{\Sigma}_k/\Sigma_k \to 1$, as $\sigma \to 0$. Using this property in (4.9) and following the same methods, one obtain the lemma by choosing an appropriate g_σ. $\square$

Lemma 4. *For all integer N, define*

$$g(N,y) = \mathbb{E}_\theta\left[\sum_{k=1}^{N}(1-a_k^2)b_k^{-2}y_k^2 + \varepsilon^2\sum_{k=1}^{N}(a_k^2-1)b_k^{-2}\right] + (1+\alpha)\mathbb{E}\{U_0(N,x) - U_0(N)\}. \tag{4.19}$$

Let $\bar{N}_0$ and $\tilde{N}_0$ defined in (3.2). *There exists $\delta_\sigma \to 0$ as $\sigma \to 0$ and $C > 0$ such that, for all $\gamma > 0$:*

$$\mathbb{E}_\theta g(\tilde{N}_0, y) \leqslant (\delta_\sigma + \gamma)\mathbb{E}_\theta[R_\alpha(\theta,\tilde{N}_0)] + \gamma R_\alpha(\theta,\bar{N}_0) + \frac{C\sigma_1^2}{\gamma^{4\beta+1}}\left(\frac{\sigma}{\varepsilon}\|\theta\|\right)^{4\beta+1} + C\varepsilon^2 + \Omega.$$

Proof. Let $\tilde{N}$ measurable w.r.t. a sequence $(\vartheta_k)_{k\in\mathbb{N}}$ of i.i.d. standard Gaussian random variables. For all $N \in \mathbb{N}$ and $\gamma > 0$, inequality (4.31) of Cavalier and Golubev (2006) provide that

$$\mathbb{E}_\theta\sum_{k=1}^{\tilde{N}}\varepsilon b_k^{-1}\theta_k\vartheta_k \leqslant \gamma\left(\sum_{k>N}\theta_k^2 + \varepsilon^2\sum_{k=1}^{N}b_k^{-2}\right) + \gamma\mathbb{E}_\theta\left(\sum_{k>\tilde{N}}\theta_k^2 + \varepsilon^2\sum_{k=1}^{\tilde{N}}b_k^{-2}\right) + \frac{C\varepsilon^2}{\gamma^{2\beta+1}}. \tag{4.20}$$

Using Lemma 5 and (4.20):

$$\mathbb{E}_\theta\left[\sum_{k=1}^{\tilde{N}_0}(1-a_k^2)b_k^{-2}y_k^2\right] \leqslant \mathbb{E}_\theta\left[\sum_{k=1}^{\tilde{N}_0}\frac{\sigma}{b_k}\zeta_k b_k^{-2}y_k^2\right] \leqslant C\left(\log^{-1/3}\frac{1}{\sigma}\right)\mathbb{E}_\theta\left[\sum_{k=1}^{\tilde{N}_0}\varepsilon^2 b_k^{-2}\right] + \mathbb{E}_\theta\left[\sum_{k=1}^{\tilde{N}_0}\frac{\sigma}{\varepsilon}\theta_k^2\varepsilon b_k^{-1}\zeta_k\right] + \Omega$$
$$\leqslant (\rho_\sigma + \gamma)\mathbb{E}_\theta[R_\alpha(\theta,\tilde{N}_0)] + \gamma R_\alpha(\theta,\bar{N}_0) + \frac{C\sigma_1^2}{\gamma^{4\beta+1}}\left(1 + \frac{\sigma}{\varepsilon}\|\theta\|\right)^{4\beta+1} + \Omega,$$

with $\rho_\sigma = \log^{-1} 1/\sigma$. Then,

$$\mathbb{E}_\theta\left[\sum_{k=1}^{\tilde N_0}(a_k^2-1)b_k^{-2}\right] \leqslant C\mathbb{E}_\theta\left[\sum_{k=1}^{\tilde N_0}\left\{\frac{\sigma}{b_k}|\zeta_k| + \frac{\sigma^2}{b_k^2}\zeta_k^2\right\}\varepsilon^2 b_k^{-2}\right] + \Omega, \leqslant \log^{-1/3}\left(\frac{1}{\sigma}\right)\mathbb{E}_\theta\left[\sum_{k=1}^{\tilde N_0}\varepsilon^2 b_k^{-2}\right] + \Omega.$$

We eventually apply Lemma 3 to complete the proof. □

The following lemma was introduced in Cavalier and Hengartner (2005).

Lemma 5. *For all $k \in \mathbb{N}$, set $a_k = x_k^{-1} b_k$ and consider the Taylor expansions*:

(i) $$a_k = 1 - \frac{\sigma}{b_k}\zeta_k + \frac{\sigma^2}{b_k^2}\zeta_k^2 v_{k,1},$$

(ii) $$a_k^2 = 1 - 2\frac{\sigma}{b_k}\zeta_k + 6\frac{\sigma^2}{b_k^2}\zeta_k^2 v_{k,2}^2,$$

where $v_{k,1}$ and $v_{k,2}$ are random variables. On the event $\mathscr{B}$, $v_{k,1}$ and $v_{k,2}$ are uniformly bounded. Moreover:

$$P(\mathscr{M}^c) \leqslant CM_0\,\mathrm{e}^{-\log^{1+\tau}1/\sigma} \quad \text{and} \quad P(\mathscr{B}^c) \leqslant CM_1\,\mathrm{e}^{-\log^{1+\tau}1/\sigma},$$

where $\mathscr{M} = \{M_0 < M < M_1\}$. The constants τ and C are positive and independent of σ.

Proof. The proof is given in Cavalier and Hengartner (2005). □

5. Proof of Theorem 1

From Theorem 2 of Cavalier and Golubev (2006), the quantity

$$V_\mu(\theta, N) = \sum_{k=N+1}^{+\infty}\theta_k^2 + \varepsilon^2\sum_{k=1}^{N} b_k^{-2} + (1+\mu)U_0(N) + \frac{C^\star\sigma_1^2}{\mu}$$

is a risk hull for all $\mu > 0$. Hence,

$$\mathbb{E}_\theta \sup_N [l(\theta, N) - V_\mu(\theta, N)] \leqslant 0.$$

In particular

$$\mathbb{E}_\theta l(\theta, N^\star) \leqslant \mathbb{E}_\theta V_\mu(\theta, N^\star). \tag{5.1}$$

Moreover, using (2.13)

$$V_\alpha(x, y, N^\star) \leqslant V_\alpha(x, y, N), \quad \forall N \leqslant M. \tag{5.2}$$

The proof of Theorem 1 consists in writing $V_\alpha(x, y, N^\star)$ in terms of $V_\mu(\theta, N^\star)$ and then using (5.1), (5.2) and Lemma 1.

Let $N^\star$ defined in (2.13), $\alpha > 1$ and $\mu > 0$.

$$\begin{aligned} V_\alpha(x,y,N^\star) - V_\mu(\theta,N^\star) &= -\sum_{k=1}^{N^\star} x_k^{-2}y_k^2 + 2\varepsilon^2\sum_{k=1}^{N^\star}x_k^{-2} + (1+\alpha)U_0(N^\star,x) - \sum_{k>N^\star}\theta_k^2 - \varepsilon^2\sum_{k=1}^{N^\star}b_k^{-2} - (1+\mu)U_0(N^\star) - \frac{C^\star\sigma_1^2}{\mu} \\ &= -\sum_{k=1}^{N^\star} x_k^{-2}y_k^2 + \varepsilon^2\sum_{k=1}^{N^\star}x_k^{-2} - \sum_{k>N^\star}\theta_k^2 - \varepsilon^2\sum_{k=1}^{N^\star}b_k^{-2}(1-a_k^2) - \frac{C^\star\sigma_1^2}{\mu} \\ &\quad + [(1+\alpha)U_0(N^\star,x) - (1+\mu)U_0(N^\star)], \end{aligned} \tag{5.3}$$

where $a_k = b_k/x_k$ for all $k \in \mathbb{N}$. Remark that

$$\begin{aligned} -\sum_{k=1}^{N^\star} x_k^{-2}y_k^2 + \varepsilon^2\sum_{k=1}^{N^\star}x_k^{-2} - \sum_{k>N^\star}\theta_k^2 &= -\varepsilon^2\sum_{k=1}^{N^\star}x_k^{-2}(\xi_k^2-1) - 2\sum_{k=1}^{N^\star}\theta_k x_k^{-2}b_k\varepsilon\xi_k - \sum_{k=1}^{N^\star}a_k^2\theta_k^2 - \sum_{k>N^\star}\theta_k^2 \\ &= -\varepsilon^2\sum_{k=1}^{N^\star}x_k^{-2}(\xi_k^2-1) - 2\varepsilon\sum_{k=1}^{N^\star}\theta_k x_k^{-2}b_k\xi_k - \|\theta\|^2 - \sum_{k=1}^{N^\star}(a_k^2-1)\theta_k^2. \end{aligned} \tag{5.4}$$

Now, we use inequality (ii) of Lemma 5 in order to bound the sums containing the terms a_k in (5.3) and (5.4). These sums represent the error made when using x_k^{-2} instead of b_k^{-2}. A string of inequalities similar to the proof of Lemma 1 reveals that

$$\mathbb{E}_\theta \sum_{k=1}^{N^\star}(a_k^2-1)\theta_k^2 \leqslant Cc_\sigma \|\theta\|^2 \frac{\sigma^2}{\varepsilon^2}\mathbb{E}_\theta \varDelta(N^\star) - 2\mathbb{E}_\theta \left[\sum_{k=1}^{N^\star}\sigma b_k^{-1}\theta_k^2\zeta_k\right] + \Omega, \tag{5.5}$$

where $c_\sigma = \log^{4/3} 1/\sigma$. Moreover,

$$\varepsilon^2 \mathbb{E}_\theta \sum_{k=1}^{N^\star} b_k^{-2}(1-a_k^2) \leqslant \kappa_\sigma \mathbb{E}_\theta V_\nu(\theta, N^\star) + \Omega, \tag{5.6}$$

where $\kappa_\sigma \to 0$ as $\sigma \to 0$. Indeed, Lemma 5 and (4.6) provide that $a_k \to 1$ uniformly in $k \leqslant M$ on the event $\mathscr{B}$. Then use (5.3)–(5.6) to obtain

$$\begin{aligned}(1-\kappa_\sigma)\mathbb{E}_\theta V_\mu(\theta,N^\star) \leqslant{}& \|\theta\|^2 + \frac{C^\star\sigma_1^2}{\mu} + Cc_\sigma \frac{\sigma^2}{\varepsilon^2}\|\theta\|^2 \mathbb{E}_\theta \varDelta(N^\star) + \mathbb{E}_\theta \Bigg[V_\alpha(x,y,N^\star) + \varepsilon^2 \sum_{k=1}^{N^\star} x_k^{-2}(\xi_k^2-1) + 2\varepsilon \sum_{k=1}^{N^\star}\theta_k a_k x_k^{-1}\xi_k \\ &- \{(1+\alpha)U_0(N^\star,x) - (1+\mu)U_0(N^\star)\} - 2\sigma\sum_{k=1}^{N^\star} b_k^{-1}\zeta_k\theta_k^2\Bigg] + \Omega. \end{aligned} \tag{5.7}$$

Remark that

$$-2\sigma \mathbb{E}_\theta \sum_{k=1}^{N^\star} b_k^{-1}\zeta_k\theta_k^2 = 2\sigma \mathbb{E}_\theta \sum_{k=1}^{N^\star} b_k^{-1}\bar{\zeta}_k\theta_k^2,$$

where $\bar{\zeta}_k \sim \mathcal{N}(0,1)$. Let $\bar{N}_0$ defined in (3.2). Using (4.20), for all $\gamma_1 > 0$,

$$\begin{aligned}-2\mathbb{E}_\theta \sum_{k=1}^{N^\star}\sigma b_k^{-1}\zeta_k\theta_k^2 &\leqslant \gamma_1\left(\sum_{k=\bar{N}_0+1}^{+\infty}\theta_k^2 + \sum_{k=1}^{\bar{N}_0}\varepsilon^2 b_k^{-2}\right) + \frac{C\sigma_1^2}{\gamma_1^{4\beta+1}}\left(1+\frac{\sigma}{\varepsilon}\|\theta\|\right)^{4\beta+1} + \gamma_1\mathbb{E}_\theta\left(\sum_{k=N^\star+1}^{+\infty}\theta_k^2 + \varepsilon^2\sum_{k=1}^{N^\star} b_k^{-2}\right) \\ &\leqslant \gamma_1 \mathbb{E}_\theta V_\mu(\theta,N^\star) + \gamma_1 R_\alpha(\theta,\bar{N}_0) + \frac{C\sigma_1^2}{\gamma_1^{4\beta+1}}\left(1+\frac{\sigma}{\varepsilon}\|\theta\|\right)^{4\beta+1}. \end{aligned} \tag{5.8}$$

Now using (5.2), $V_\alpha(x,y,N^\star) \leqslant V_\alpha(x,y,N)$ for all $N \leqslant M$. In particular, this inequality is true for $N = \tilde{N}_0$ where

$$\tilde{N}_0 = arg \inf_{N \leqslant M} R_\alpha(\theta, N), \tag{5.9}$$

and $R_\alpha(\theta,N)$ is defined in (3.1) for all $N \in \mathbb{N}$. The bandwidth $\tilde{N}_0$ is stochastic. It depends on the sequence $(x_k)_{k\geqslant 1}$. Nevertheless, since $(x_k)_{k\in\mathbb{N}}$ and $(y_k)_{k\in\mathbb{N}}$ are independent,

$$\begin{aligned}\mathbb{E}_\theta V_\alpha(x,y,\tilde{N}_0) &= -\mathbb{E}_\theta\left[\sum_{k=1}^{\tilde{N}_0} x_k^{-2}y_k^2\right] + 2\varepsilon^2 \mathbb{E}_\theta\left[\sum_{k=1}^{\tilde{N}_0} x_k^{-2}\right] + (1+\alpha)\mathbb{E}_\theta U_0(\tilde{N}_0,x) \\ &= \mathbb{E}_\theta\left[-\sum_{k=1}^{\tilde{N}_0}\theta_k^2 + \sum_{k=1}^{\tilde{N}_0} b_k^{-2} + (1+\alpha)U_0(\tilde{N}_0)\right] + \mathbb{E}_\theta[g(\tilde{N}_0,y)] \\ &= -\|\theta\|^2 + \mathbb{E}_\theta R_\alpha(\theta,\tilde{N}_0) + \mathbb{E}_\theta g(\tilde{N}_0,y), \end{aligned} \tag{5.10}$$

where, for all $N \in \mathbb{N}$,

$$g(N,y) = \mathbb{E}_\theta\left[\sum_{k=1}^{N}\left(1-a_k^2\right)b_k^{-2}y_k^2 + \varepsilon^2\sum_{k=1}^{N}\left(a_k^2-1\right)b_k^{-2}\right] + (1+\alpha)\,\mathbb{E}_\theta\left\{U_0(N,x) - U_0(N)\right\}. \tag{5.11}$$

Using inequalities (5.7)–(5.11), we obtain, for all $N \leqslant M$:

$$\begin{aligned}(1-\gamma_1-\kappa_\sigma)\mathbb{E}_\theta V_\mu(\theta,N^\star) \leqslant{}& \mathbb{E}_\theta[R_\alpha(\theta,\tilde{N}_0)] + \varepsilon^2\mathbb{E}_\theta\left[\sum_{k=1}^{N^\star} x_k^{-2}\left(\xi_k^2-1\right)\right] + 2\mathbb{E}_\theta\left[\sum_{k=1}^{N^\star}\theta_k a_k \varepsilon x_k^{-1}\xi_k\right] \\ &+ \frac{C\sigma_1^2}{\gamma_1^{4\beta+1}}\left(1+\frac{\sigma}{\varepsilon}\|\theta\|\right)^{4\beta+1} - \mathbb{E}_\theta\left\{(1+\alpha)U_0(N^\star,x) - (1+\mu)U_0(N^\star)\right\} \\ &+ Cc_\sigma\frac{\sigma^2}{\varepsilon^2}\|\theta\|^2\mathbb{E}_\theta\left[\varDelta(N^\star)\right] + \mathbb{E}_\theta\left[g(\tilde{N}_0,y)\right] + \gamma_1 R_\alpha(\theta,\bar{N}_0) + \frac{C^\star\sigma_1^2}{\mu} + \Omega. \end{aligned}$$

Then apply Lemmas 2, 4 and (4.20). For all $\gamma_2>\gamma_1>0$ and $0<B<1$,

$$(1-\gamma_1-B-\kappa_\sigma)\mathbb{E}_\theta V_\mu(\theta,N^\star)\leqslant(1+2\gamma_2+\delta_\sigma)\mathbb{E}_\theta R_\alpha(\theta,\tilde{N}_0)+\frac{C\sigma_1^2}{\gamma_1^{4\beta+1}}\left(1+\frac{\sigma}{\varepsilon}\|\theta\|\right)^{4\beta+1}+\varepsilon^2(1-\gamma_2)\mathbb{E}_\theta\sum_{k=1}^{N^\star}x_k^{-2}(\xi_k^2-1)$$
$$-\mathbb{E}_\theta\{(1+\alpha)U_0(N^\star,x)-(1+\mu)U_0(N^\star)\}+\gamma_2\mathbb{E}_\theta l(\theta,N^\star)+\varepsilon^2B^{-2\beta}\left(\frac{\sigma^2}{\varepsilon^2}\|\theta\|^2\right)^{2\beta+1}+2\gamma_1R_\alpha(\theta,\bar{N}_0)+\frac{C^\star\sigma_1^2}{\mu}+\Omega. \quad (5.12)$$

Now consider the term

$$T=\varepsilon^2(1-\gamma_2)\mathbb{E}_\theta\sum_{k=1}^{N^\star}x_k^{-2}(\xi_k^2-1)-\mathbb{E}_\theta\left\{(1+\alpha)U_0(N^\star,x)-(1+\mu)U_0(N^\star)\right\}.$$

This represents the main difficulty of the proof. Indeed, inequality (4.24) of Cavalier and Golubev (2006) establishes that for all $v>0$,

$$\mathbb{E}_\theta\sup_k\{\eta_k-(1+v)U_0(k)\}\leqslant\frac{C\sigma_1^2}{v}.$$

This result can be generalized to the quantity T, conditioning on the sequence $(x_k)_{k\geqslant1}$. Lemma 3 provides that there exists g_σ verifying

$$\mathbb{E}_\theta\sup_{N\leqslant M}[U_0(N)-(1+g_\sigma)U_0(N,x)]\leqslant C\varepsilon^2+\Omega$$

for some $C>0$. Conditioning on the sequence $(x_k)_{k\geqslant1}$, we obtain the following bound:

$$T\leqslant(1-\gamma_2)\mathbb{E}_\theta\left[\tilde{\eta}_{N^\star}-\frac{\alpha-\mu-(1+\mu)g_\sigma}{1-\gamma_2}U_0(N^\star,x)\right]+C\varepsilon^2+\Omega\leqslant\frac{(1-\gamma_2)^2C\sigma_1^2}{(\alpha-\mu(1+g_\sigma)-g_\sigma+\gamma_2-1)_+}+C\varepsilon^2+\Omega. \quad (5.13)$$

If the term g_σ is negative, we can easily find μ and γ_2 able to control the residual term in (5.13). The choice is the same as in Cavalier and Golubev (2006). Now, if g_σ is positive and cannot be controlled, any acceptable bound for (5.13) is available. The noisy penalty $U_0(N^\star,x)$ is too small to contain the stochastic term in $V_\alpha(x,y,N^\star)$. Fortunately, Lemma 3 ensures that $g_\sigma<1$. Choose

$$1>\gamma_2>g_\sigma. \quad (5.14)$$

Since $\alpha-1>0$, set

$$\mu=\frac{\alpha-1}{2(1+g_\sigma)}.$$

With this choice of parameters, there exists C independent of ε such that $T\leqslant C\varepsilon^2/(\alpha-1)_+$. Now, setting $B=\gamma_1$, use (5.1) and (5.13) to obtain

$$(1-\gamma_2-2\gamma_1-\kappa_\sigma)\mathbb{E}_\theta l(\theta,N^\star)\leqslant(1+2\gamma_2+\delta_\sigma)\mathbb{E}_\theta R_\alpha(\theta,\tilde{N}_0)+2\gamma_1R_\alpha(\theta,\bar{N}_0)+\frac{C\sigma_1^2}{\gamma_1^{4\beta+1}}\left(1+\frac{\sigma^2}{\varepsilon^2}\|\theta\|^2\right)^{2\beta+1}+\frac{C\varepsilon^2}{(\alpha-1)}+\Omega,$$

since α is greater then 1. Whatever the value of γ_2, the parameter γ_1 can be chosen sufficiently small in order to provide that

$$1-\gamma_2-2\gamma_1-\kappa_\sigma>0,$$

since for a fixed $C>0$, κ_σ can be construced so that $\kappa_\sigma<C$. Then remark that

$$\mathbb{E}_\theta R_\alpha(\theta,\tilde{N}_0)=\mathbb{E}_\theta\inf_{N\leqslant M}R_\alpha(\theta,N)\leqslant\inf_N R_\alpha(\theta,N)+\Gamma(\theta)+\Omega.$$

We eventually apply Lemmas 1 and 2 with $B=\gamma_1$ to obtain

$$\mathbb{E}_\theta\|\theta^\star-\theta\|^2\leqslant(1+\rho_\sigma)\frac{(1+\delta_\sigma+2\gamma_2)}{(1-2\gamma_1-\gamma_2-\kappa_\sigma)}R_\alpha(\theta,\bar{N}_0)+\frac{C\varepsilon^2}{(\alpha-1)_+}+\frac{C\sigma_1^2}{\gamma_1^{2\beta+1}}\left(1+\frac{\sigma^2}{\varepsilon^2}\|\theta\|^2\right)^{2\beta+1}+\Gamma(\theta)+\Omega$$

for some $1>\gamma>0$, where ρ_σ is defined in Lemma 1. In order to conclude the proof, set for example $\gamma_2=\sqrt{g_\sigma}$ in (5.14). $\square$

References

Barron, A., Birgé, L., Massart, P., 1999. Risk bounds for model selection via penalization. Probab. Theory Related Fields 113, 301–413.
Bertero, M., Bindi, D., Boccacci, P., Cattaneo, M., Eva, C., Lanza, V., 1998. A novel blind-deconvolution method with an application to seismology. Inverse Problems 14, 815–833.
Birgé, L., Massart, P., 2001. Gaussian model selection. J. European Math. Soc. 3, 203–268.
Cavalier, L., Golubev, Y., 2006. Risk hull method and regularization by projections of ill-posed inverse problems. Ann. Statist. 34, 1653–1677.
Cavalier, L., Hengartner, N.W., 2005. Adaptive estimation for inverse problems with noisy operators. Inverse Problems 21, 1345–1361.
Cavalier, L., Raimondo, M., 2005. Wavelet deconvolution with noisy eigenvalues. IEEE Trans. Signal Process. 55, 2414–2424.
Cavalier, L., Golubev, G.K., Picard, D., Tsybakov, A.B., 2002. Oracle inequalities for inverse problems. Ann. Statist. 30, 843–874.
Efromovich, S., Koltchinskii, V., 2001. On inverse problems with unknown operators. IEEE Trans. Inform. Theory 47, 2876–2893.
Engl, H., Hanke, M., Neubauer, A., 1996. Regularization of Inverse Problems. Kluwer Academic Publishers Group, Dordrecht.
Fan, J., 1991. On the optimal rates of convergence for nonparametric deconvolution problems. Ann. Statist. 19, 1257–1272.
Golubev, Y., 2004. The principle of penalized empirical risk in severely ill-posed problems. Theory Probab. Appl. 130, 18–38.
Hida, T., 1980. Brownian Motion. Springer, New York, Berlin.
Hoffmann, M., Reiß, M., 2008. Nonlinear estimation for linear inverse problems with error in the operator. Ann. Statist. 38, 310–336.
Johnstone, I.M., Silverman, S.W., 1990. Speed of estimation in positron emission tomography and related inverse problems. Ann. Statist. 18, 251–280.
Johnstone, I.M., Kerkyacharian, G., Picard, D., Raimondo, R., 2004. Wavelet deconvolution in a periodic setting. J. Roy. Statist. Soc. B 66 (Part 3), 547–573.
Korostelev, A.P., Tsybakov, A.B., 1993. Minimax theory of image reconstruction. In: Lecture Notes in Statistics, vol. 82, Springer, New York.
Loubes, J.M., Ludena, C., 2007. Penalized estimators for non-linear inverse problems, in preparation.
Loubes, J.M., Fermin, A.K., Ludena, C., 2005. Model selection for ill-posed inverse problems. In: Proceedings of Oberwolfach.
Marteau, C., 2006. Regularization of inverse problems with unknown operator. Math. Methods Statist. 15, 415–443.
Pruessner, A., O'Leary, D., 2003. Blind deconvolution using a regularized structured total least norm algorithm. SIAM J. Matrix Anal. Appl. 24, 1018–1037.