HAL open science

Quadratic functional estimation in inverse problems

Cristina Butucea, Katia Meziani

▶ To cite this version:

Cristina Butucea, Katia Meziani. Quadratic functional estimation in inverse problems. Statistical Methodology, 2010, 8 (1), pp.31-41. 10.1016/j.stamet.2010.05.002 . hal-00634945

HAL Id: hal-00634945

https://hal.science/hal-00634945

Submitted on 24 Oct 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Quadratic functional estimation in inverse problems

Cristina BUTUCEA[1], Katia MEZIANI[2]
[1] Laboratoire Paul Painlevé (UMR CNRS 8524),
Université des Sciences et Technologies de Lille 1,
59655 Villeneuve d'Ascq cedex, France
e-mail: cristina.butucea@math.univ-lille1.fr
[2]Laboratoire de Probabilités et Modèles Aléatoires,
Université Paris VII (Denis Diderot),
75251 Paris Cedex 05, France
email: meziani@math.jussieu.fr

February 10, 2009

Abstract

We consider in this paper a Gaussian sequence model of observations Y_i, $i \geq 1$ having mean (or signal) θ_i and variance σ_i which is growing polynomially like i^γ, $\gamma > 0$. This model describes a large panel of inverse problems. We estimate the quadratic functional of the unknown signal $\sum_{i\geq 1} \theta_i^2$ when the signal belongs to ellipsoids of both finite smoothness functions (polynomial weights i^α, $\alpha > 0$) and infinite smoothness (exponential weights $e^{\beta i^r}$, $\beta > 0$, $0 < r \leq 2$). We propose a Pinsker type projection estimator in each case and study its quadratic risk. When the signal is sufficiently smoother than the difficulty of the inverse problem ($\alpha > \gamma + 1/4$ or in the case of exponential weights), we obtain the parametric rate and the efficiency constant associated to it. Moreover, we give upper bounds of the second order term in the risk and conjecture that they are asymptotically sharp minimax. When the signal is finitely smooth with $\alpha \leq \gamma + 1/4$, we compute non parametric upper bounds of the risk of and we presume also that the constant is asymptotically sharp.

Mathematics Subject Classifications 2000: 62F12, 62G05, 62G10, 62G20

Key Words: Gaussian sequence model, inverse problem, minimax upper bounds, parametric rate, Pinsker estimator, projection estimator, quadratic functional, second order risk.

1 Introduction

We observe $\{Y_i\}_{i=1\cdots n}$

$$Y_i = \theta_i + \epsilon\,\xi_i \quad \forall i = 1\cdots n \tag{1}$$

where ξ_i are independent identically distributed (i.i.d.) random variables, having a Gaussian law with zero mean and variance $\sigma_i^2 = i^{2\gamma}$ for some fixed $\gamma \geq 0$. Let us mention that in case $\{\sigma_i\}_{i\geq 1}$ is a bounded sequence the problem is direct and when $\sigma_i \to \infty$ the problem is an inverse problem. We say that the problem is ill-posed when σ_i increases polynomially and severely ill-posed when it increases exponentially.

We want to estimate the quadratic functional $Q(\theta) = \sum_{i=1}^{\infty} \theta_i^2$, where $\theta = \{\theta_i\}_{i\geq 1}$ belongs to the ℓ_2-ellipsoid

$$\Sigma = \left\{ \theta : \sum_{i=1}^{\infty} a_i^2 \theta_i^2 \leq L \right\}, \tag{2}$$

where a_i is a non decreasing sequence of positive real numbers and $L > 0$. We consider both polynomial sequence $a_i = i^{\alpha}$ where we say that the signal is (ordinary) smooth and exponential sequence $a_i = \exp(\beta i^r)$ where we say that the signal is super-smooth, α, $\beta > 0$ and $0 < r \leq 2$.

It is known that this model can be deduced from a linear operator equation with noisy observations $Y = Ax + \epsilon\,\xi$, where $A : \mathcal{H} \to \mathcal{H}$ is a known linear operator on the Hilbert space $\mathcal{H}$, x belongs to $\mathcal{H}$ is the signal of interest and ξ is a standard white Gaussian noise. By considering an orthonormal basis $\{\varphi_i\}_{i\geq 1}$ of $\mathcal{H}$, we consider only the sequence of values $Y_i := Y(\varphi_i)/b_i$, where b_i^2 are the eigenvalues of AA^* for $i \geq 1$. For more details and examples of inverse problems that can be written in the form (1) we refer the reader to Cavalier *et al.*[4], [5] and references therein. We mention as particular examples the convolution operator, the Radon transform in the case of tomography or problems described by partial differential equations.

Estimation of θ in the inverse problem (1) with a quadratic risk was thoroughly studied in the literature from the minimax point of view. Let us only mention a few minimax adaptive results: oracle inequalities in Cavalier *et al.* [5], sharp adaptive estimation by block thresholding in Cavalier *et al.* [4] and adaptive estimators defined by penalized empirical risk in Golubev [9] .

Estimation of quadratic functionals in inverse problems was studied in two particular problems (specified operators). Butucea [1] considered the convolution density model and studied the rates of a kernel type estimator. Méziani [14] estimates the purity of a quantum state, which corresponds mathematically to a quadratic functional of a bivariate function of mass 1, in a double inverse problem: tomography and convolu-

tion with Gaussian noise. Our model allows to consider more general inverse problems, i.e. various operators A.

Quadratic functionals were much more studied in the direct problem (σ_j bounded for all j) since first results given by Ibragimov and Has'minskiĭ [10] and Ibragimov *et al.* [11]. Fan [8] gave minimax rates over hyperrectangles and Sobolev-type ellipsoids. Donoho and Nussbaum [6] gave Pinsker sharp minimax estimators in this model and in the equivalent models of fixed equidistant design regression and Gaussian white noise model. Fore more general bodies which are not quadratically convex, Cai and Low [2] showed that nonquadratic estimators attain the minimax rate of the quadratic functional. For adaptive estimators over hyperrectangles we cite Efromovich and Low [7]. Sharp or nearly sharp adaptive estimators over l_p-bodies were found by Klemelä [12]. Adaptive estimators over more general Besov and l_p bodies were given by Cai and Low [3]. In the density model, let us mention adaptive estimators via model selection by Laurent [13].

Let us underline the difference between estimating $Q(\theta)$ in our model and that of estimating from direct data $\sum_{j\geq 1} j^{2\gamma}\theta_j^2$ for $\gamma \in \mathbb{N}$ as it was done, e.g., by Fan [8], Donoho and Nussbaum [6] and Klemelä [12]. In our case, the variance of our estimators is slower. When estimating the quadratic functional of a derivative, the bias is smaller, so the rates and constants are different.

Here, we give a Pinsker-type projection estimator which automatically attains the parametric rate and the efficiency constant for all super-smooth signals and for the smooth signals when $\alpha \geq \gamma + 1/4$. Moreover, in this case we give nonparametric minimax upper bounds of the second order term in the quadratic risk. Our estimator attains the expected minimax nonparametric rate in the case of smooth signals with $\alpha < \gamma + 1/4$. We conjecture that the asymptotic constant in the nonparametric upper bound of the risk is sharp. The proofs of sharp lower bounds will make the object of future work.

Let us mention that our method can be easily adapted for severely ill posed inverse problems, i.e. σ_i increases as an exponential. The case where $\sigma_i = e^{i^2}$ is of particular interest in practice and hasn't been studied for estimating the signal $\{\theta_i\}_{i\geq 1}$ either. Future developments should concern adaptive estimation of the quadratic functional.

In Section 2 we describe the estimator and the precise choice of tuning parameters and give asymptotic upper bounds rates of convergence and associated constant. We postpone the proofs to the Section 3 and the Appendix.

2 Estimation procedure and results

Let us define the estimator

$$\widetilde{Q} = \sum_{i=1}^{\infty} h_i (Y_i^2 - \epsilon^2 \sigma_i^2), \tag{3}$$

where $\{h_i\}_{i\geq 1}$ is a sequence between 0 and 1. We shall actually see that the optimal sequence is truncated, i.e. $h_i = 0$ for all $i > W$ and that the optimal value of W tends to infinity when $\epsilon \to 0$.

Let us first consider the case of smooth signal: $\theta \in \Sigma(\alpha, L)$, where $a_i = i^\alpha$.

Theorem 1 *Let observations $Y_1, \ldots, Y_n, \ldots$ satisfy model (1). Then the estimator $\widetilde{Q}$ in (3) with parameters $\{h_i\}_{i\geq 1}$ and W defined by*

$$\begin{aligned} h_i &= \left(1 - \left(\frac{i}{W}\right)^{2\alpha}\right)_+ \quad \textit{and} \\ W &= \left\lfloor \left(\frac{L^2(4\gamma + 4\alpha + 1)(4\gamma + 2\alpha + 1)}{4\alpha}\right)^{\frac{1}{4\alpha+4\gamma+1}} \epsilon^{-\frac{4}{4\alpha+4\gamma+1}} \right\rfloor \end{aligned}$$

is such that

$$\sup_{\theta\in\Sigma(\alpha,L)} E\left[\left(\widetilde{Q} - Q(\theta)\right)^2\right] = C(\alpha,\, \gamma,\, L)\epsilon^{\frac{16\alpha}{4\alpha+4\gamma+1}}(1 + o(1)),$$

if $\alpha \leq \gamma + \frac{1}{4}$,

$$\sup_{\theta\in\Sigma(\alpha,L)} E\left[\left(\widetilde{Q} - Q(\theta)\right)^2 - 4\epsilon^2 \sum_{i=1}^{\infty} \sigma_i^2 \theta_i^2\right] = C(\alpha,\, \gamma,\, L)\epsilon^{\frac{16\alpha}{4\alpha+4\gamma+1}}(1 + o(1)),$$

if $\alpha > \gamma + \frac{1}{4}$, where

$$C(\alpha,\, \gamma,\, L) = \frac{L^{2\frac{4\gamma+1}{4\alpha+4\gamma+1}}}{(4\gamma+1)} \left(\frac{2\alpha + 4\gamma + 1}{4\alpha}\right)^{-\frac{4\alpha}{4\alpha+4\gamma+1}} (4\alpha + 4\gamma + 1)^{\frac{4\gamma+1}{4\alpha+4\gamma+1}}. \tag{4}$$

We find a known phenomenon in quadratic functional estimation literature, i.e. the existence of two cases: a regular one, where the rate is parametric ϵ^{-2}, and an irregular case when the rate is significantly slower. We conjecture that Theorem 1 exhibits sharp asymptotic constant in this last case.

In the regular case (when the underlying signal is smoother than the 'difficulty' of the operator A), Theorem 1 says actually two things. One of them is that, for each θ in the set Σ the quadratic risk of our estimator is of parametric rate and attains the efficiency constant in our model:

$$E\left[\left(\widetilde{Q} - Q(\theta)\right)^2\right] = 4\epsilon^2 \sum_{i=1}^{\infty} \sigma_i^2 \theta_i^2 (1 + o(1)),$$

as $\epsilon \to 0$. Secondly, the quadratic risk is decomposed and the second order risk is optimized for our choice of parameters and equals the risk in the non parametric case.

Note also, that the rates are not surprising when compared to the results of Butucea [1] for the convolution density model. No second order terms were evaluated there, nor constants associated to the nonparametric rate. The efficiency constant is naturally different for the density model.

Let us now consider the case of super-smooth signal: $\theta \in \Sigma(\beta, r, L)$, where $a_i = \exp(\beta i^r)$.

Theorem 2 *Let observations $Y_1, \ldots, Y_n, \ldots$ satisfy model (1). Let the estimator $\widetilde{Q}$ in (3) be defined with parameters $\{h_i\}_{i\geq 1}$ given by*

$$h_i = \left(1 - \frac{e^{2\beta i^r}}{e^{2\beta W^r}}\right)_+$$

and W solution of the equation

$$W^{4\gamma+(1-r)_+} \exp(4\beta W^r - 2\beta r W^{r-1} I_{(r>1)}) = c(\beta, r, \gamma, L)\epsilon^{-4},$$

with the constant $c := c(\beta, r, \gamma, L) = 2\beta r L^2$ if $0 < r < 1$, $c = L^2(e^{4\beta} - 1)/(2e^{2\beta})$ if $r = 1$, $c = L^2/2$ if $1 < r < 2$ and $c = L^2/(2e^{2\beta})$ if $r = 2$. Then

$$\sup_{\theta\in\Sigma(\beta,r,L)} E\left[\left(\widetilde{Q} - Q(\theta)\right)^2 - 4\epsilon^2 \sum_{i=1}^{\infty} \sigma_i^2 \theta_i^2\right] = \frac{2\epsilon^4}{4\gamma+1}\left(\frac{\log(1/\epsilon)}{\beta}\right)^{(4\gamma+1)/r}(1 + o(1)).$$

We note that in this case, the signal is always smoother than the difficulty of the inverse problem, so there is always a parametric rate term in the quadratic risk. Our estimator also optimizes the upper bounds for the second order term in the quadratic risk. In this last term, the bias term is always smaller than the variance term for super-smooth signals.

3 Proofs

Proof of Theorem 1. We decompose as usually the quadratic risk $E\left[\left(\widetilde{Q} - Q(\theta)\right)^2\right]$ into bias plus variance. The bias term can be written

$$\begin{aligned}\left(E[\widetilde{Q}] - Q(\theta)\right)^2 &= \left(\sum_{i=1}^{\infty} h_i E[Y_i^2 - \epsilon^2\sigma_i^2] - \sum_{i=1}^{\infty}\theta_i^2\right)^2 \\ &= \left(\sum_{i=1}^{\infty} h_i\theta_i^2 - \sum_{i=1}^{\infty}\theta_i^2\right)^2 = \left(\sum_{i=1}^{\infty}\theta_i^2(1-h_i)\right)^2. \qquad (5)\end{aligned}$$

The variance term is decomposed as follows

$$\begin{aligned} E\left[\left(\widetilde{Q}-E[\widetilde{Q}]\right)^2\right] &= E\left[\left(\sum_{i=1}^{\infty} h_i(Y_i^2-\epsilon^2\sigma_i^2)-\sum_{i=1}^{\infty} h_i\theta_i^2\right)^2\right] \\ &= E\left[\left(\sum_{i=1}^{\infty} h_i(Y_i^2-\epsilon^2\sigma_i^2-\theta_i^2)\right)^2\right]. \end{aligned}$$

Since Y_i are independent and ξ_i are independent Gaussian random variables:

$$\begin{aligned} E\left[\left(\widetilde{Q}-E[\widetilde{Q}]\right)^2\right] &= \sum_{i=1}^{\infty} h_i^2 E\left[(Y_i^2-\epsilon^2\sigma_i^2-\theta_i^2)^2\right] \\ &= \sum_{i=1}^{\infty} h_i^2 E\left[(2\epsilon\theta_i\xi_i-\epsilon^2\sigma_i^2+\epsilon^2\xi_i^2)^2\right] \\ &= \sum_{i=1}^{\infty} h_i^2 \left\{\epsilon^4 E\left[\xi_i^4\right]-2\epsilon^4\sigma_i^2 E\left[\xi_i^2\right]+4\epsilon^2\theta_i^2 E\left[\xi_i^2\right]+\epsilon^4\sigma_i^4\right\}. \end{aligned} \tag{6}$$

Now, use the facts that $E[\xi_i^2]=\sigma_i^2$ and $E[\xi_i^4]=3\sigma_i^4$ to get

$$\begin{aligned} E\left[\left(\widetilde{Q}-E[\widetilde{Q}]\right)^2\right] &= 4\epsilon^2\sum_{i=1}^{\infty} h_i^2\sigma_i^2\theta_i^2+2\epsilon^4\sum_{i=1}^{\infty} h_i^2\sigma_i^4 \\ &= 4\epsilon^2\sum_{i=1}^{\infty}\sigma_i^2\theta_i^2-4\epsilon^2\sum_{i=1}^{\infty}(1-h_i^2)\sigma_i^2\theta_i^2+2\epsilon^4\sum_{i=1}^{\infty} h_i^2\sigma_i^4 \end{aligned} \tag{7}$$

Thus by (5) and (7) we get

$$E\left[\left(\widetilde{Q}-Q(\theta)\right)^2\right]=A_0(h,\theta)+A_1(h)+A_2(\theta)-A_3(h,\theta), \tag{8}$$

where

$$\begin{aligned} A_0(h,\theta)=A_0 &:= \left(\sum_{i=1}^{\infty}\theta_i^2(1-h_i)\right)^2, \\ A_1(h)=A_1 &:= 2\epsilon^4\sum_{i=1}^{\infty} h_i^2\sigma_i^4, \\ A_2(\theta)=A_2 &:= 4\epsilon^2\sum_{i=1}^{\infty}\sigma_i^2\theta_i^2, \\ A_3(h,\theta)=A_3 &:= 4\epsilon^2\sum_{i=1}^{\infty}(1-h_i^2)\sigma_i^2\theta_i^2. \end{aligned}$$

If we note $T(h,\theta):=A_0(h,\theta)+A_1(h)=A_0+A_1$, then we want to find

$$\inf_h \sup_{\theta\in\Sigma} T(h,\theta)\le \sup_{\theta\in\Sigma} T(h,\theta)\le \sup_{\theta\in\partial\Sigma} T(h,\theta)$$

where the infimum is taken with respect to all sequences h such that $0 \leq h_i \leq 1$ for all $i \geq 1$ and with

$$\partial\Sigma = \left\{ \theta : \sum_{i=1}^{\infty} a_i^2 \theta_i^2 = L \right\}. \tag{9}$$

Let us define $F(h,\theta) = T(h,\theta) - \kappa \left(\sum_{i=1}^{\infty} a_i^2 \theta_i^2 - L \right)$ with $\kappa > 0$. Then for all $j \in \mathbb{N}^*$ the optimal h and θ have to verify

$$\frac{\partial}{\partial \theta_j} F(h,\theta) = 0 \quad \text{and} \quad \frac{\partial}{\partial h_j} F(h,\theta) = 0.$$

We get

$$\begin{aligned} h_j &= \left(1 - \frac{\kappa a_j^2}{2 \sum_{i=1}^{\infty} \theta_i^2 (1-h_i)} \right)_+ = \left(1 - \widetilde{\kappa} a_j^2 \right)_+, \\ (\theta_j^*)^2 &= \frac{2\epsilon^4 \sigma_j^4 h_j}{\sum_{i=1}^{\infty} \theta_i^2 (1-h_i)}, \end{aligned} \tag{10}$$

where $\widetilde{\kappa} > 0$. Let us write $h_j = \left(1 - \frac{j^{2\alpha}}{W^{2\alpha}} \right)_+$ where $W \to \infty$ when $\epsilon \to 0$.

Recall that $\Sigma = \Sigma(\alpha, L) = \left\{ \theta : \sum_{i=1}^{\infty} i^{2\alpha} \theta_i^2 \leq L \right\}$ then for $\theta^* \in \partial\Sigma(\alpha, L)$ we can write both

$$\begin{aligned} \sum_{i=1}^{\infty} \theta_i^{*2} (1-h_i) &= \frac{1}{W^{2\alpha}} \sum_{i=1}^{W} \theta_i^{*2} i^{2\alpha} + \sum_{i>W} \theta_i^{*2} \\ &\leq \frac{1}{W^{2\alpha}} \sum_{i=1}^{W} \theta_i^{*2} i^{2\alpha} + \frac{1}{W^{2\alpha}} \sum_{i>W} \theta_i^{*2} i^{2\alpha} \leq \frac{L}{W^{2\alpha}} \end{aligned}$$

and

$$\begin{aligned} \sum_{i=1}^{\infty} \theta_i^{*2} (1-h_i) &\geq \frac{1}{W^{2\alpha}} \sum_{i=1}^{W} \theta_i^{*2} i^{2\alpha} \\ &\geq \frac{1}{W^{2\alpha}} \sum_{i=1}^{\infty} \theta_i^{*2} i^{2\alpha} - \frac{1}{W^{2\alpha}} \sum_{i>W} \theta_i^{*2} i^{2\alpha} = \frac{L}{W^{2\alpha}} (1 - o(1)). \end{aligned}$$

Therefore $A_0 = L^2 W^{-4\alpha} (1 + o(1))$, as $\epsilon \to 0$. This means also that we can write

$$(\theta_j^*)^2 = \frac{2\epsilon^4 \sigma_j^4 W^{2\alpha}}{L} \left(1 - \frac{j^{2\alpha}}{W^{2\alpha}} \right)_+.$$

Let us now compute the optimal W, using again the fact that $\theta^* \in \partial\Sigma(\alpha, L)$ which is equivalent to

$$\sum_{i=1}^{\infty} i^{2\alpha} (\theta_i^*)^2 = L.$$

This is further equivalent to

$$W^{4\alpha+4\gamma+1}\frac{1}{W}\sum_{i=1}^{W}\left(\frac{i}{W}\right)^{4\gamma+2\alpha}\left(1-\left(\frac{i}{W}\right)^{2\alpha}\right)=\frac{L^2}{2\epsilon^4}$$

giving

$$\frac{2\alpha W^{4\alpha+4\gamma+1}}{(4\gamma+4\alpha+1)(4\gamma+2\alpha+1)}(1+o(1))=\frac{L^2}{2\epsilon^4}.$$

Therefore

$$W=\left(\frac{L^2}{B(\alpha,\gamma)}\right)^{\frac{1}{4\alpha+4\gamma+1}}\epsilon^{-\frac{4}{4\alpha+4\gamma+1}}(1+o(1)), \tag{11}$$

where $B(\alpha,\gamma):=\frac{4\alpha}{(4\gamma+4\alpha+1)(4\gamma+2\alpha+1)}$ and we'll take W to be the integer part of the dominant term. From now on, we denote $B:=B(\alpha,\gamma)$.

We have to evaluate the term defined in (8). For $\alpha\leq\gamma+\frac{1}{4}$, we have

$$\begin{aligned}
A_0 &= \left(\sum_{i=1}^{\infty}\theta_i^2(1-h_i)\right)^2=L^2W^{-4\alpha}(1+o(1))\\
&= \left(L^{2(4\gamma+1)}B^{4\alpha}\right)^{\frac{1}{4\gamma+4\alpha+1}}\epsilon^{\frac{16\alpha}{4\gamma+4\alpha+1}}(1+o(1)),\\
A_1 &= 2\epsilon^4\sum_{i=1}^{\infty}\sigma_i^4h_i^2=2\epsilon^4W^{4\gamma+1}\frac{1}{W}\sum_{i=1}^{W}\left(\frac{i}{W}\right)^{4\gamma}\left(1-\left(\frac{i}{W}\right)^{2\alpha}\right)^2\\
&= \frac{16\alpha^2\epsilon^4W^{4\gamma+1}}{(4\gamma+1)(4\gamma+4\alpha+1)(4\gamma+2\alpha+1)}(1+o(1))\\
&= \frac{4\alpha}{4\gamma+1}\left(L^{2(4\gamma+1)}B^{4\alpha}\right)^{\frac{1}{4\gamma+4\alpha+1}}\epsilon^{\frac{16\alpha}{4\gamma+4\alpha+1}}(1+o(1)),\\
A_2 &= 4\epsilon^2\sum_{i=1}^{\infty}\sigma_i^2\theta_i^{*2}=\frac{8\epsilon^6W^{6\gamma+2\alpha+1}}{L}\frac{1}{W}\sum_{i=1}^{W}\left(\frac{i}{W}\right)^{6\gamma}\left(1-\left(\frac{i}{W}\right)^{2\alpha}\right)\\
&= \frac{16\alpha\epsilon^6W^{6\gamma+2\alpha+1}}{L(6\gamma+1)(6\gamma+2\alpha+1)}(1+o(1))\\
&= O(1)\epsilon^{\frac{16\alpha+2}{4\alpha+4\gamma+1}}(1+o(1))=o(1)A_1,
\end{aligned}$$

as $\epsilon\to0$. As $h_i\in[0,1]$ for all $i\in\mathbb{N}$, the term $A_3=4\epsilon^2\sum_{i=1}^{\infty}(1-h_i^2)\sigma_i^2\theta_i^2\leq A_2$. Then the quadratic risk is such that

$$\begin{aligned}
E\left[\left(\widetilde{Q}-Q(\theta)\right)^2\right] &= (A_0+A_1)(1+o(1))\\
&= \left(L^{2(4\gamma+1)}B^{4\alpha}\right)^{\frac{1}{4\gamma+4\alpha+1}}\frac{4\gamma+4\alpha+1}{4\gamma+1}\epsilon^{\frac{16\alpha}{4\gamma+4\alpha+1}}(1+o(1)),
\end{aligned}$$

as $\epsilon\to0$ and this explains the constant $C(\alpha,\gamma,L)$ in (4).

Let us note that if $\alpha > \gamma + \frac{1}{4}$, we can estimate the quadratic functional at the parametric rate as A_2 is the dominant term in the risk and is of order ϵ^2. More precisely

$$E\left[\left(\widetilde{Q} - Q(\theta)\right)^2\right] = 4\epsilon^2 \sum_{i=1}^{\infty} \sigma_i^2 \theta_i^2 \left(1 + o(1)\right) = A_2(1 + o(1)),$$

as $\epsilon \to 0$. Indeed, it is easy to see that in this case

$$A_0 + A_1 = C(\alpha, \gamma, L)\epsilon^{\frac{16\alpha}{4\gamma+4\alpha+1}}(1 + o(1)) = o(A_2)$$

and, moreover,

$$\begin{aligned} A_3 &= 4\epsilon^2 \sum_{i=1}^{W} \left[1 - \left(1 - \frac{i^{2\alpha}}{W^{2\alpha}}\right)^2\right] i^{2\gamma}\theta_i^2 + 4\epsilon^2 \sum_{i>W} i^{2\gamma}\theta_i^2 \\ &\leq 4\epsilon^2 \sum_{i=1}^{W} \frac{i^{2\alpha+2\gamma}}{W^{2\alpha}} + 4\epsilon^2 \sum_{i>W} i^{2(\gamma-\alpha)} i^{2\alpha}\theta_i^2 \\ &\leq 4\epsilon^2 W^{2(\gamma-\alpha)} \sum_{i=1}^{W} \left(\frac{i}{W}\right)^{2\gamma} i^{2\alpha}\theta_i^2 + 4\epsilon^2 W^{2(\gamma-\alpha)} \sum_{i=1}^{W} i^{2\alpha}\theta_i^2 \\ &\leq 4\epsilon^2 W^{2(\gamma-\alpha)} L = O(1)\epsilon^{\frac{16\alpha+2}{4\alpha+4\gamma+1}} = o(A_0 + A_1), \end{aligned}$$

as $\epsilon \to 0$. ■

Proof of Theorem 2. We follow the lines of proof of Theorem 1. In this case, there is always a parametric term and we do the computations of the second order term in the quadratic risk.

We solve the same optimisation problem and find

$$h_i = \left(1 - \frac{e^{2\beta i^r}}{e^{2\beta W^r}}\right)_+ \qquad (\theta_j^*)^2 = \frac{2\epsilon^4 \sigma_j^4 h_j}{\sum_{i=1}^{\infty} \theta_i^2 (1 - h_i)}. \tag{12}$$

Then for $\theta^* \in \partial\Sigma(\beta, L, r)$ we get

$$\begin{aligned} \sum_{i=1}^{\infty} \theta_i^{*2}(1 - h_i) &= \frac{1}{e^{2\beta W^r}} \sum_{i=1}^{W} e^{2\beta i^r}\theta_i^{*2} + \sum_{i>W} \theta_i^{*2} \\ &\leq \frac{1}{e^{2\beta W^r}} \sum_{i=1}^{W} e^{2\beta i^r}\theta_i^{*2} + \frac{1}{e^{2\beta W^r}} \sum_{i>W} e^{2\beta i^r}\theta_i^{*2} = \frac{L}{e^{2\beta W^r}} \end{aligned}$$

and

$$\begin{aligned} \sum_{i=1}^{\infty} \theta_i^{*2}(1 - h_i) &\geq \frac{1}{e^{2\beta W^r}} \sum_{i=1}^{W} e^{2\beta i^r}\theta_i^{*2} \\ &= \frac{1}{e^{2\beta W^r}} \sum_{i=1}^{\infty} e^{2\beta i^r}\theta_i^{*2} - \frac{1}{e^{2\beta W^r}} \sum_{i>W} e^{2\beta i^r}\theta_i^{*2} = \frac{L}{e^{2\beta W^r}}(1 - o(1)). \end{aligned}$$

Therefore

$$A_0 = L^2 e^{-4\beta W^r}(1+o(1)), \text{ as } \epsilon \to 0.$$

By (12), this gives $\theta_i^{*2} = \frac{2\epsilon^4\sigma_j^4}{L}\left(e^{2\beta W^r} - e^{2\beta j^r}\right)_+$.

To compute optimal W, we also use the fact $\theta^* \in \partial\Sigma(\beta, L, r)$.

$$\sum_{i=1}^{\infty} e^{2\beta i^r}(\theta_i^*)^2 = L \Leftrightarrow e^{2\beta W^r}\sum_{i=1}^{W-1} i^{4\gamma}e^{2\beta i^r} - \sum_{i=1}^{W-1} i^{4\gamma}e^{4\beta i^r} = \frac{L^2}{2\epsilon^4}$$

By using Lemmata 1 and 2, we have W solution of the following equation

$$\begin{aligned} &W^{4\gamma}e^{4\beta W^r - 2\beta r W^{r-1}} = c\epsilon^{-4}, && \text{if } 1 < r \le 2, \\ &W^{4\gamma}e^{4\beta W} = c\epsilon^{-4}, && \text{if } r = 1, \\ &W^{4\gamma - r + 1}e^{4\beta W^r} = c\epsilon^{-4}, && \text{if } 0 < r < 1, \end{aligned} \tag{13}$$

as $\epsilon \to 0$, with the constant $c = c(\beta, \gamma, L)$ defined in Theorem 2.

We evaluate $A_0 + A_1$: in each of the previous cases, the bias term A_0 is infinitely smaller than the variance term A_1 and the main term in A_1 can be given for

$$W = \left(\frac{\log(1/\epsilon)}{\beta}\right)^{1/r}.$$

Indeed, by using Lemmata 1 and 2,

$$\begin{aligned} A_1 &= 2\epsilon^4\sum_{i=1}^{\infty}\sigma_i^4 h_i^2 = 2\epsilon^4\sum_{i=1}^{W} i^{4\gamma}\left(1 - \frac{e^{2\beta i^r}}{e^{2\beta W^r}}\right)^2 \\ &= \frac{2\epsilon^4 W^{4\gamma+1}}{4\gamma+1}(1+o(1)) = \frac{2\epsilon^4}{4\gamma+1}\left(\frac{\log(1/\epsilon)}{\beta}\right)^{(4\gamma+1)/r}(1+o(1)) = o(A_2). \end{aligned}$$

As $A_0 = o(A_1)$ it is easy to see that in this case

$$A_0 + A_1 = \frac{2\epsilon^4}{4\gamma+1}\left(\frac{\log(1/\epsilon)}{\beta}\right)^{(4\gamma+1)/r}(1+o(1)) = o(A_2)$$

as $\epsilon \to 0$.

The last thing to check is that $A_3 = o(A_0 + A_1)$ as $\epsilon \to 0$:

$$\begin{aligned} A_3 &= 4\epsilon^2\sum_{i=1}^{\infty}(1-h_i^2)\sigma_i^2\theta_i^{*2} \le 8\epsilon^2\sum_{i=1}^{W}\frac{e^{2\beta i^r}}{e^{2\beta W^r}}i^{2\gamma}\theta_i^{*2} + 4\epsilon^2\sum_{i>W} i^{2\gamma}\theta_i^{*2} \\ &\le 8\epsilon^2\frac{W^{2\gamma}}{e^{2\beta W^r}}\sum_{i=1}^{W} e^{2\beta i^r}\theta_i^{*2} + 8\epsilon^2\sum_{i>W} i^{2\gamma}\frac{e^{2\beta i^r}}{e^{2\beta i^r}}\theta_i^{*2} \\ &\le 8\epsilon^2\frac{W^{2\gamma}}{e^{2\beta W^r}}\sum_{i=1}^{W} e^{2\beta i^r}\theta_i^{*2} + 8\epsilon^2\frac{W^{2\gamma}}{e^{2\beta W^r}}\sum_{i>W} e^{2\beta i^r}\theta_i^{*2} \\ &= 8\epsilon^2\frac{W^{2\gamma}}{e^{2\beta W^r}}L = O(1)W^{4\gamma+1}\epsilon^4\frac{1}{W^{2\gamma+1}\epsilon^2 e^{2\beta W^r}}. \end{aligned}$$

So, we can write that

$$A_3 = O(A_1)\frac{1}{W^{2\gamma+1}\epsilon^2 e^{2\beta W^r}}.$$

By (13), we easily see that

$$\begin{aligned}
&W^{2\gamma+1}e^{2\beta W^r-\beta r W^{r-1}}\epsilon^2 = \sqrt{c}\,W, && \text{if } 1<r\leq 2,\\
&W^{2\gamma+1}e^{2\beta W}\epsilon^2 = \sqrt{c}\,W, && \text{if } r=1,\\
&W^{2\gamma+1}e^{2\beta W}\epsilon^2 = \sqrt{c}\,W^{(1+r)/2}, && \text{if } 0<r<1,
\end{aligned}$$

Then, as $W\to\infty$, we get for all $r\in]0,2]$, $A_3 = o(A_1)$ as $\epsilon\to 0$.

■

4 Appendix

Lemma 1 *For all a, b, $s>0$ and $v>0$*

$$\int_0^v x^a e^{bx^s}dx \;=\; \frac{v^{a-s+1}e^{bv^s}}{bs}(1+o(1)),$$

as $v\to\infty$.

Lemma 2 *For $a\geq 0$, $b>0$, and $r>0$ as $N\to\infty$*

$$\sum_{i=1}^N i^a e^{bi^r} = \begin{cases} N^a e^{bN^r}(1+o(1)) \text{ if } r>1,\\ \frac{1}{br}N^{a+1-r}e^{bN^r}(1+o(1)) \text{ if } 0<r<1,\\ \frac{1}{(e^b-1)}N^a e^{b(N+1)}(1+o(1)) \text{ if } r\text{=1 and } a\neq 0.\end{cases}$$

Proof of Lemma 2. • When $r>1$

$$\begin{aligned}
\sum_{i=1}^N i^a e^{bi^r} - N^a e^{bN^r} &= \sum_{i=1}^{N-1} i^a e^{bi^r} \leq (N-1)^{a+1}e^{b(N-1)^r}\\
&\leq N^a e^{bN^r}O(N)e^{-brN^{r-1}} = o(1)N^a e^{bN^r},
\end{aligned}$$

as $N\to\infty$.

• When $0<r<1$

$$\int_1^{N+1} x^a e^{bx^r}dx \geq \sum_{i=1}^N i^a e^{bi^r} \geq \int_0^N x^a e^{bx^r}dx.$$

Use Lemma 1 and the fact that

$$\int_1^{N+1} x^a e^{bx^r}dx = \int_0^N x^a e^{bx^r}dx(1+o(1).$$

• When $r=1$ we write both

$$\sum_{i=1}^N i^a e^{bi} \;=\; N^a e^{bN} + \sum_{i=1}^{N-1} i^a e^{bi}$$

and

$$\sum_{i=1}^{N} i^a e^{bi} \quad = \quad e^b \sum_{i=0}^{N-1} (i+1)^a e^{bi} = e^b + e^b \sum_{i=1}^{N-1} (i+1)^a e^{bi}.$$

As the sums $\sum_{i=1}^{N-1} i^a e^{bi}$ and $\sum_{i=1}^{N-1} (i+1)^a e^{bi}$ have equivalent general terms and diverge, than they are equivalent to S_{N-1}, say. We get that, for large N,

$$S_N = \frac{N^a e^{b(N+1)}}{e^b - 1}(1 + o(1)).$$

■

References

[1] Butucea, C. Goodness-of-fit testing and quadratic functional estimation from indirect observations. *Ann. Statist. 35*, 5 (2007), 1907–1930.

[2] Cai, T. T., and Low, M. G. Nonquadratic estimators of a quadratic functional. *Ann. Statist. 33*, 6 (2005), 2930–2956.

[3] Cai, T. T., and Low, M. G. Optimal adaptive estimation of a quadratic functional. *Ann. Statist. 34*, 5 (2006), 2298–2325.

[4] Cavalier, L., Golubev, G. K., Picard, D., and Tsybakov, A. B. Oracle inequalities for inverse problems. *Ann. Statist. 30*, 3 (2002), 843–874. Dedicated to the memory of Lucien Le Cam.

[5] Cavalier, L., Golubev, Y., Lepski, O., and Tsybakov, A. Block thresholding and sharp adaptive estimation in severely ill-posed inverse problems. *Teor. Veroyatnost. i Primenen. 48*, 3 (2003), 534–556.

[6] Donoho, D. L., and Nussbaum, M. Minimax quadratic estimation of a quadratic functional. *J. Complexity 6*, 3 (1990), 290–323.

[7] Efromovich, S., and Low, M. On optimal adaptive estimation of a quadratic functional. *Ann. Statist. 24*, 3 (1996), 1106–1125.

[8] Fan, J. On the estimation of quadratic functionals. *Ann. Statist. 19* (1991), 1273–1294.

[9] Golubev, Y. The principle of penalized empirical risk in severely ill-posed problems. *Probab. Theory Related Fields 130*, 1 (2004), 18–38.

[10] Has′minskiĭ, R. Z., and Ibragimov, I. A. Some estimation problems for stochastic differential equations. In *Stochastic differential systems (Proc. IFIP-WG 7/1 Working Conf., Vilnius, 1978)*, vol. 25 of *Lecture Notes in Control and Information Sci.* Springer, Berlin, 1980, pp. 1–12.

[11] Ibragimov, I. A., Nemirovskiĭ, A. S., and Khas′minskiĭ, R. Z. Some problems of nonparametric estimation in Gaussian white noise. *Teor. Veroyatnost. i Primenen. 31*, 3 (1986), 451–466.

[12] Klemelä, J. Sharp adaptive estimation of quadratic functionals. *Probab. Theory Related Fields 134*, 4 (2006), 539–564.

[13] Laurent, B. Adaptive estimation of a quadratic functional of a density by model selection. *ESAIM Probab. Stat. 9* (2005), 1–18 (electronic).

[14] Méziani, K. Nonparametric estimation of the purity of a quantum state in quantum homodyne tomography with noisy data. *Math. Methods Statist. 16*, 4 (2007), 354–368.