ISSN 1066-5307, Mathematical Methods of Statistics, 2015, Vol. 24, No. 1, pp. 1–15. © Allerton Press, Inc., 2015.

Estimation in Ill-posed Linear Models with Nuisance Design

Yu. Golubev[1*] **and T. Zimolo**[2**]

[1] *Aix-Marseille Univ., Marseille, France;*
Inst. Inform. Transm. Probl., Moscow, Russia

[2] *Aix-Marseille Univ., Marseille, France*

Received October 29, 2014; in final form, February 06, 2015

Abstract—The paper deals with recovering an unknown vector $\theta \in \mathbb{R}^p$ in two simple linear models: in the first one we observe $y = b \cdot \theta + \epsilon\xi$ and $z = b + \sigma\xi'$, whereas in the second one we have at our disposal $y' = b^2 \cdot \theta + \epsilon b \cdot \xi$ and $z = b + \sigma\xi'$. Here $b \in \mathbb{R}^p$ is a nuisance vector with positive components and ξ, $\xi' \in \mathbb{R}^p$ are standard white Gaussian noises in $\mathbb{R}^p$. It is assumed that p is large and components b_k of b are small for large k. In order to get good statistical estimates of θ in this situation, we propose to combine minimax estimates of $1/b_k$ and $1/b_k^2$ with regularization techniques based on the roughness penalty approach. We provide new non-asymptotic upper bounds for the mean square risks of the estimates related to this method.

Keywords: noisy deconvolution, inverse minimax estimation, Van Trees inequality, roughness penalty approach.

2000 Mathematics Subject Classification: Primary 62C99; secondary 62C10, 62C20, 62J05.

DOI: 10.3103/S1066530715010019

1. INTRODUCTION

This paper deals with estimating an unknown vector $\theta \in \mathbb{R}^p$ in two simple linear models. In the first one θ is estimated based on the data

$$\begin{aligned} y_k &= b_k\theta_k + \epsilon\xi_k, \quad k = 1, \dots, p, \\ z_k &= b_k + \sigma\xi_k', \quad k = 1, \dots, p, \end{aligned} \tag{1}$$

whereas in the second one θ is recovered from the observations

$$\begin{aligned} y_k' &= b_k^2\theta_k + \epsilon b_k\xi_k, \quad k = 1, \dots, p, \\ z_k &= b_k + \sigma\xi_k', \quad k = 1, \dots, p, \end{aligned} \tag{2}$$

where ξ and ξ' are independent standard white Gaussian noises in $\mathbb{R}^p$ and $b \in \mathbb{R}^p$ is an unknown nuisance vector with nonnegative components $b_k \geq 0$, $k = 1, 2, \dots, p$. In order to simplify numerous technical details, it is assumed in what follows that the noise levels ϵ an σ are known.

In spite of very simple probabilistic structures of (1) and (2), estimation of θ in these statistical models is a nontrivial problem. Principal difficulties arise when

$$p \text{ is large} \quad \text{and} \quad b_k \text{ are small.}$$

The basic idea to overcome these difficulties is based on regularization methods which nowadays are well developed in the case $\sigma = 0$. These methods are usually related to the roughness penalty approach and the main goal in this paper is to adapt this approach to the case $\sigma > 0$.

*E-mail: golubev.yuri@gmail.com
**E-mail: thomas.zimolo@univ-amu.fr

Linear models (1) and (2) play an important role in studying, for instance, *the noisy periodic deconvolution problem*. Suppose we have at our disposal the noisy data

$$\begin{aligned} Y(t) &= \int_0^1 h(t-u)X(u)\,du + \epsilon n(t), \quad t \in [0,1], \\ Z(t) &= h(t) + \sigma n'(t), \quad t \in [0,1], \end{aligned} \tag{3}$$

where

- $n(t)$ and $n'(t)$ are independent standard white Gaussian noises;
- $h(t),\ t \in [0,1]$, is an unknown periodic function with period 1.

Our goal is to recover $X(t),\ t \in [0,1]$, based on the observation $\{Y(t),\ Z(t),\ t \in [0,1]\}$.

The continuous time model (3) can be easily transformed into the so-called sequence space model with the help of the standard trigonometric basis on $[0,1]$

$$\varphi_0(t) = 1,\ \varphi_k(t) = \sqrt{2}\cos(2\pi kt),\ \varphi_k^*(t) = \sqrt{2}\sin(2\pi kt), \quad k = 1, 2, \dots.$$

Denote for brevity

$$\begin{aligned} X_0 &= \int_0^1 X(t)\varphi_0(t)\,dt,\ X_k = \int_0^1 X(t)\varphi_k(t)\,dt,\ X_k^* = \int_0^1 X(t)\varphi_k^*(t)\,dt; \\ Y_0 &= \int_0^1 Y(t)\varphi_0(t)\,dt,\ Y_k = \int_0^1 Y(t)\varphi_k(t)\,dt,\ Y_k^* = \int_0^1 Y(t)\varphi_k^*(t)\,dt; \\ Z_0 &= \int_0^1 Z(t)\varphi_0(t)\,dt,\ Z_k = \int_0^1 Z(t)\varphi_k(t)\,dt,\ Z_k^* = \int_0^1 Z(t)\varphi_k^*(t)\,dt; \\ h_0 &= \int_0^1 h(t)\varphi_0(t)\,dt,\ h_k = \int_0^1 h(t)\varphi_k(t)\,dt,\ h_k^* = \int_0^1 h(t)\varphi_k^*(t)\,dt. \end{aligned}$$

Then with a simple algebra we arrive at the following statistical model:

$$\begin{aligned} &Y_0 = h_0 X_0 + \epsilon\xi_0, \\ &Y_k = \frac{X_k h_k - X_k^* h_k^*}{\sqrt{2}} + \epsilon\xi_k, \qquad Y_k^* = \frac{X_k h_k^* + X_k^* h_k}{\sqrt{2}} + \epsilon\xi_{-k}; \\ &Z_0 = h_0 + \sigma\xi_0', \\ &Z_k = h_k + \sigma\xi_k', \qquad Z_k^* = h_k^* + \sigma\xi_{-k}', \end{aligned} \tag{4}$$

which is equivalent to (3). In the above equations, ξ and ξ' are independent white Gaussian noises.

Suppose $h(\cdot)$ is a symmetric function with $h_k > 0$. This means that $h_k^* = 0$. In other words, we assume that the convolution operator $H\colon L_2(0,1) \to L_2(0,1)$ defined by

$$Hx(t) = \int_0^1 h(t-u)x(u)\,du, \quad t \in [0,1],$$

is self-adjoint and positively defined. In this case, estimation of X_k, $k = 0, 1, \dots$, in (4) is equivalent to estimation of X_k based on the data

$$\begin{aligned} Y_0 &= h_0 X_0 + \epsilon\xi_0, \qquad & Y_k &= \frac{X_k h_k}{\sqrt{2}} + \epsilon\xi_k; \\ Z_0 &= h_0 + \sigma\xi_0', \qquad & Z_k &= h_k + \sigma\xi_k' \end{aligned}$$

and to estimation of $X_k^*,\ k = 1, 2, \dots$, with the help of the observations

$$Y_k^* = \frac{X_k^* h_k}{\sqrt{2}} + \epsilon\xi_{-k}; \qquad Z_k = h_k + \sigma\xi_k'.$$

Thus we see that if H is *a self-adjoint and positively defined operator*, then the noisy deconvolution is equivalent to recovering $\theta \in l_2$ in Model (1).

In the general case, one can rewrite (4) in the following equivalent form:

$$
\begin{aligned}
&Y_0 = h_0 X_0 + \epsilon\xi_0,\\
&Y_k Z_k + Y_k^* Z_k^* = \frac{X_k(h_k^2 + h_k^{*2})}{\sqrt{2}}\\
&\qquad + \epsilon(\xi_k h_k + \xi_{-k} h_k^*) + \epsilon\sigma(\xi_{-k}\xi_k' + \xi_{-k}\xi_{-k}'),\\
&Y_k^* Z_k - Y_k^* Z_k^* = \frac{X_k^*(h_k^2 + h_k^{*2})}{\sqrt{2}}\\
&\qquad + \epsilon(\xi_{-k} h_k - \xi_k h_k^*) + \epsilon\sigma(\xi_{-k}\xi_{-k}' - \xi_k\xi_{-k}'),\\
&Z_0 = h_0 + \sigma\xi_0',\\
&Z_k^2 = h_k^2 + 2\sigma h_k \xi_k' + \sigma^2\xi_k'^2, \qquad Z_k^{*2} = h_k^{*2} + 2\sigma h_k^*\xi_{-k}' + \sigma^2\xi_{-k}'^2.
\end{aligned}
\tag{5}
$$

Therefore, denoting for brevity

$$
b_i = \sqrt{h_k^2 + h_k^{*2}}, \qquad \bar Y_k = Y_k Z_k + Y_k^* Z_k^*,
$$
$$
\bar Y_k^* = Y_k^* Z_k - Y_k^* Z_k^*, \qquad \bar Z_k = \sqrt{Z_k^2 + Z_k^{*2}}
$$

and omitting the second order terms proportional to $\sigma\epsilon$ and σ^2, we arrive at the following approximation of (5):

$$
\begin{aligned}
&Y_0 = h_0 X_0 + \epsilon\xi_0, \qquad \bar Y_k \approx \frac{X_k b_k^2}{\sqrt{2}} + \epsilon b_k \bar\xi_k, \qquad \bar Y_k^* \approx \frac{X_k^* b_k^2}{\sqrt{2}} + \epsilon b_k \bar\xi_k^*;\\
&Z_0 = h_0 + \sigma\xi_0', \qquad \bar Z_k = b_k + \sigma\bar\xi_k',
\end{aligned}
$$

where $\xi_0\, \bar\xi_k$, $\bar\xi_k^*$, $\bar\xi_k'$ are mutually independent standard Gaussian random variables. So, we see that recovering X_k and X_k^* in (4) is nearly equivalent to estimating $\theta \in l_2$ in Model (2).

Another example, where statistical models similar to (1) and (2) appear, is related to the probability density deconvolution problem. Suppose we observe n i.i.d. pairs of random variables

$$
(Y_i, Z_i), \quad i = 1, \dots, n, \quad \text{where} \quad Y_i = Z_i' + X_i.
$$

The random vectors $(X_1, \dots, X_n)^\top$, $(Z_1, \dots, Z_n)^\top$ and $(Z_1', \dots, Z_n')^\top$ are assumed to be independent and the variables Z_i and Z_i' are identically distributed. The goal is to estimate the probability density of X_1. Notice also that statistical problems close to the one mentioned above are common in econometric applications related to the instrumental variables, see for instance [7], [3] and references herein.

The problem of estimation of θ in (1) has been already addressed in several papers, see for instance [1], [2], [4], [8], [6]. The principal idea in these papers is to estimate unknown b_i^{-1} using a "natural" estimate $1/z_i$ and then to correct obvious drawbacks of this method with a thresholding method.

In fact, as we will see below, estimating $1/b_i$ is a nontrivial and interesting statistical problem from a mathematical viewpoint. For instance, we can prove at the moment the optimality of proposed estimators only with the help of computerized calculations.

2. MAIN RESULTS

2.1. Univariate Minimax Inversion

The main idea in estimating $\theta \in \mathbb{R}^p$ in (1) and (2) is based on a solution to the following simple statistical problem. Suppose we observe a Gaussian random variable

$$
z = \mu + \sigma\xi, \tag{6}
$$

where $\mu \in \mathbb{R}^+$ is an unknown parameter and ξ is a standard Gaussian random variable. Our goal is to estimate $1/\mu$. More precisely, we are looking for the so-called minimax estimator $\bar{\mu}^{-1}(z)$ of $1/\mu$ and its minimax risk defined by

$$r_1(\sigma) \stackrel{\text{def}}{=} \inf_{\tilde{\mu}^{-1}} \sup_{\mu>0} \mu^4 \mathbf{E}_\mu \big[\tilde{\mu}^{-1}(z) - \mu^{-1}\big]^2 = \sup_{\mu>0} \mu^4 \mathbf{E}_\mu \big[\bar{\mu}^{-1}(z) - \mu^{-1}\big]^2, \tag{7}$$

where the infimum is taken over all measurable functions $\tilde{\mu}^{-1}(\cdot)\colon \mathbb{R}^1 \to \mathbb{R}^+$, and $\mathbf{E}_\mu$ stands for the expectation w.r.t. the probability measure generated by the observation (6).

Notice that the considered problem is closely related to estimating θ in Models (1) and (2) when $\epsilon = 0$.

We begin with a lower bound for the minimax risk $r_1(\sigma)$.

Lemma 1.

$$r_1(\sigma) \ge \sigma^2. \tag{8}$$

Proof. Inequality (8) may be proved with the help of the Van Trees inequality [10] (see also, e.g., [5]) which bounds from below the Bayesian risk of any estimate of $g(\mu)$ based on the observation $z \in \mathbb{R}^1$ with a probability density $P(\cdot;\mu)$, where $\mu \in [a,b]$ is an unknown parameter. Recall that the Bayesian risk is defined by

$$R(\pi, P) = \inf_{\tilde{g}} \int_a^b \int_{\mathbb{R}} \pi(\mu) P(z;\mu) \big[\tilde{g}(z) - g(\mu)\big]^2 \, d\mu \, dz.$$

Suppose $g(\mu)$, $\mu \in [a,b]$, is differentiable and $\pi(\cdot)$ is a probability density on $[a,b]$ such that $\pi(a) = \pi(b) = \pi'(a) = \pi'(b) = 0$ with

$$\int_a^b \frac{\pi'^2(\mu)}{\pi(\mu)} \, d\mu < \infty.$$

Then

$$R(\pi, P) \ge \frac{1}{I(\pi) + I(P)} \left[\int_a^b g'_\mu(\mu) \pi(\mu) \, d\mu\right]^2, \tag{9}$$

where Fisher's informations $I(\pi)$ and $I(P)$ are defined as

$$I(\pi) = \int_a^b \frac{\pi'^2_\mu(\mu)}{\pi(\mu)} \, d\mu \qquad \text{and} \qquad I(P) = \int_a^b \pi(\mu) \int_{\mathbb{R}} \frac{P'^2_\mu(z;\mu)}{P(z;\mu)} \, dz \, d\mu.$$

In the considered statistical problem

$$g(\mu) = \frac{1}{\mu} \qquad \text{and} \qquad P(z;\mu) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(z-\mu)^2}{2\sigma^2}\right].$$

Let us take

$$\pi(\mu) = \frac{1}{b-a} \pi_\circ \left[\frac{1}{b-a}\left(\mu - \frac{a+b}{2}\right)\right],$$

where

$$\pi_\circ(x) = 2\cos^2(\pi x), \quad x \in [-1/2, 1/2].$$

Then

$$\int_{-1/2}^{1/2} \frac{\pi'^2_\circ(x)}{\pi_\circ(x)} \, dx = 8\pi^2 \int_{-1/2}^{1/2} \sin^2(\pi x) \, dx = 4\pi^2$$

and therefore

$$I(\pi) = \frac{4\pi^2}{(b-a)^3}. \tag{10}$$

Next, we obviously have

$$\left(\int_a^b \pi(\mu) g'_\mu(\mu)\, d\mu\right)^2 = \left(\int_a^b \frac{\pi(\mu)}{\mu^2}\, d\mu\right)^2 \geq \frac{1}{b^4}. \tag{11}$$

It is also well known that

$$I(P) = \frac{1}{\sigma^2}.$$

Thus, substituting this equation and (10)–(11) in (9), we obtain

$$R(\pi, P) \geq \frac{b^{-4}}{\sigma^{-2} + 4\pi^2(b-a)^{-3}}$$

and combining this inequality with

$$r_1(\sigma) \geq a^4 R(\pi, P)$$

we arrive at

$$r_1(\sigma) \geq \frac{a^4 b^{-4}}{\sigma^{-2} + 4\pi^2(b-a)^{-3}}. \tag{12}$$

In order to finish the proof, choose $b = a + \sqrt{a}$ and take the limit in (12) as $a \to \infty$. □

Lemma 1 motivates the following definition.

Definition 1. An estimator $\bar{\mu}^{-1}(z)$ of $1/\mu$ is called strong-minimax if the following relations hold true:

- $$\sup_{\mu>0} \mu^2 \mathbf{E}_\mu \big[\bar{\mu}^{-1}(z)\mu - 1\big]^2 = \sigma^2; \tag{13}$$
- $$\sup_{\mu>0} \mathbf{E}_\mu [\bar{\mu}^{-1}(z)\mu]^2 = 1. \tag{14}$$

In order to demonstrate that strong-minimax estimators of $1/\mu$ exist, consider the following family of non-linear estimates

$$\bar{\mu}_\beta^{-1}(z) = \frac{z_+}{z^2 + \beta\sigma^2}, \quad \beta > 0, \tag{15}$$

where $z_+ = \max(z, 0)$.

There are simple heuristic arguments which help to understand where these estimates come from. Assume that the unknown parameter μ in (6) belongs to $\mathbb{R}$. As above, our goal is to estimate $1/\mu$ based on Z. Consider the following Bayesian risk:

$$R_\pi(\bar{\mu}) = \int_{-\infty}^{\infty} \pi(\mu)\mu^2 \mathbf{E}_\mu [\mu\bar{\mu}(z) - 1]^2\, d\mu,$$

where $\bar{\mu}(z)$ is an estimate of $1/\mu$ and $\pi(\cdot)$ is an a priory distribution density of μ. It can be checked with the standard arguments that

$$\arg\min_{\bar{\mu}} R_\pi(\bar{\mu}) = \int_{-\infty}^{\infty} \mu^3 \pi(\mu) \exp\left[-\frac{(z-\mu)^2}{2\sigma^2}\right] d\mu \bigg/ \int_{-\infty}^{\infty} \mu^4 \pi(\mu) \exp\left[-\frac{(z-\mu)^2}{2\sigma^2}\right] d\mu.$$

Assume that $\pi(\cdot)$ is the Cauchy density

$$\pi(\mu) = \pi_\gamma(\mu) = \frac{1}{\pi\gamma[1 + (\mu/\gamma)^2]}$$

with the scale parameter $\gamma > 0$. Then it is clear that as $\gamma \to 0$

$$\arg\min_{\bar{\mu}} R_{\pi_\gamma}(\bar{\mu}) \to \frac{z}{z^2 + \sigma^2}.$$

Unfortunately, this estimate is not minimax, but its minimax modification is given by (15), where $\beta > 1$ is a tuning parameter to be chosen properly. More precisely, for $\bar{\mu}_\beta^{-1}(z)$ the following fact holds.

Lemma 2. *There exist constants $\beta_\circ \geq 3/2$ and $\beta^\circ \leq \sqrt{7}+4$ such that $\bar{\mu}_\beta^{-1}(z)$ as in (15) is a strong-minimax estimator for any $\beta \in [\beta_\circ, \beta^\circ]$.*

Proof. Let

$$\Psi_{\xi,\beta}(x) \stackrel{\mathrm{def}}{=} \frac{[1+x\xi]_+}{(1+x\xi)^2+\beta x^2}, \quad x \in \mathbb{R}^+,$$

where ξ is a standard Gaussian random variable. Then Equations (14) and (13) are equivalent to the following ones:

$$\mathbf{E}\Psi^2_{\xi,\beta}(x) \leq 1, \quad x \geq 0, \tag{16}$$

$$\mathbf{E}[1-\Psi_{\xi,\beta}(x)]^2 \leq x^2, \quad x \geq 0. \tag{17}$$

Notice that if $x \geq 1/2\beta$, then

$$\Psi_{\xi,\beta}(x) \leq 1.$$

Indeed, the above condition is equivalent to

$$1+x\xi \leq (1+x\xi)^2+\beta x^2,$$

i.e,

$$0 \leq \beta+\xi^2+\frac{\xi}{x} = \left(\xi+\frac{1}{2x}\right)^2+\beta-\frac{1}{4x^2}.$$

So, to prove (16) is remains to verify that

$$\mathbf{E}\Psi^2_{\xi,\beta}(x) \leq 1 \quad \text{for all } x \in \left[0, \frac{1}{2\beta}\right].$$

It can be checked with a simple algebra that

$$\begin{aligned}1-\Psi_{\xi,\beta}(x) =& x\xi + x^2(\beta-\xi^2) - 3x^3\xi(\xi^2+\beta) - x^4(\xi^2+\beta)^2 \\ &+ [1-\Psi_{\xi,\beta}(x)][2x\xi+x^2(\xi^2+\beta)]^2.\end{aligned} \tag{18}$$

We begin with lower and upper bounds for β. Notice that for small x we have from (18)

$$1-\Psi_{\xi,\beta}(x) = x\xi + O(x^2)$$

and so

$$1-\Psi_{\xi,\beta}(x) = x\xi + x^2(\beta-\xi^2) + x^3\xi(\xi^2-3\beta) + O(x^4).$$

Therefore

$$\mathbf{E}\Psi^2_{\xi,\beta}(x) = 1 + x^2(3-2\beta) + O(x^4) \tag{19}$$

and

$$\mathbf{E}\big[\Psi_{\xi,\beta}(x)-1\big]^2 = x^2 + x^4(\beta^2-8\beta+9) + O(x^6). \tag{20}$$

Hence with (19) we obtain that $\beta_\circ \geq 3/2$. On the other hand, with (20) we arrive at

$$\beta_\circ > 4-\sqrt{7} \approx 1.35 \qquad \text{and} \qquad \beta^\circ \leq 4+\sqrt{7} \approx 6.65.$$

In order to obtain more precise bounds for β, we computed numerically the following functions:

$$r_0(\beta) = \sup_{x\geq 0} \mathbf{E}\Psi^2_{\xi,\beta}(x) \qquad \text{and} \qquad r_1(\beta) = \sup_{x \geq 0} x^{-2}\mathbf{E}\big[\Psi_{\xi,\beta}(x)-1\big]^2.$$

Their plots are shown in Fig. 1. We see that $3/2$ is the exact lower bound for β, i.e., $\beta_\circ = 1.5$, whereas $\beta^\circ \approx 2.7$. □

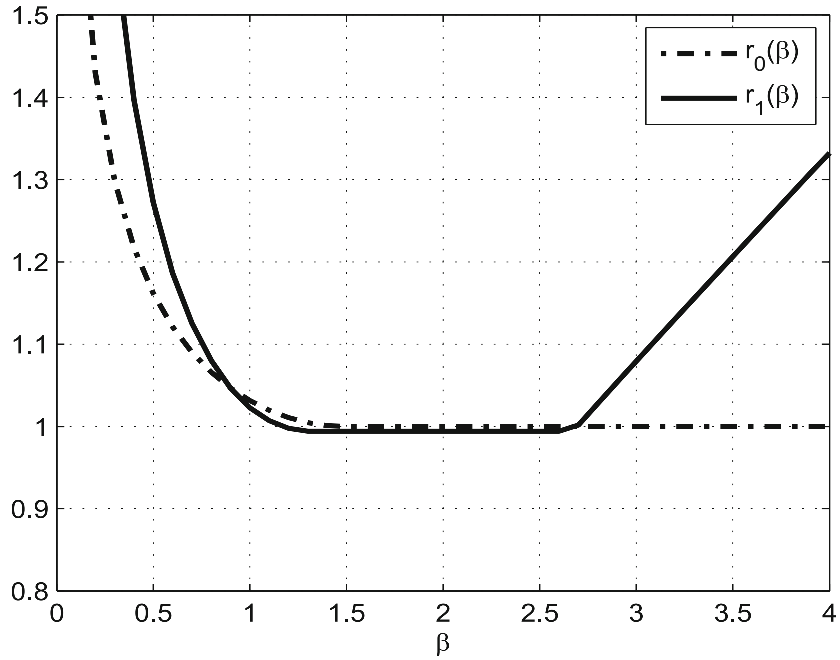

Fig. 1. Risk functions $r_0(\beta)$ and $r_1(\beta)$ for $\bar{b}^{-1}(z)$.

In fact, the family of strong-minimax estimators of $1/\mu$ is wide. For instance, along with the Bayesian approach, such estimators can be obtained by the roughness penalty method. A simple example of such an estimator is given by

$$\bar{\mu}_\beta(z) = \arg\max_{\mu>0}\left\{-\frac{(z-\mu)^2}{2\sigma^2} + \beta\log(\mu)\right\} = \frac{z}{2} + \sqrt{\frac{z^2}{4} + \beta\sigma^2}. \tag{21}$$

With this estimate we arrive at the following estimate of $1/\mu$:

$$\tilde{\mu}_\beta^{-1}(z) = \frac{1}{\bar{\mu}_\beta(z)} = \frac{1}{\beta\sigma^2}\left[\sqrt{\frac{z^2}{4} + \beta\sigma^2} - \frac{z}{2}\right]. \tag{22}$$

For this method a fact similar to Lemma 2 holds.

Lemma 3. *There exist constants $\tilde{\beta}_\circ$, $\tilde{\beta}^\circ$ such that $\tilde{\mu}_\beta^{-1}(z)$ is strong-minimax for any $\beta \in [\tilde{\beta}_\circ, \tilde{\beta}^\circ]$.*

At the moment we cannot provide an analytical proof of this result. The computerized proof is based on computing the risk functions

$$r_0(\beta) = \sup_{\mu>0}\mathbf{E}_\mu[\mu\tilde{\mu}^{-1}(z)]^2 \qquad \text{and} \qquad r_1(\beta) = \sup_{\mu>0}\mu^2\mathbf{E}_\mu[\mu\tilde{\mu}^{-1}(z) - 1]^2$$

shown in Fig. 2.

Comparing Figs. 1 and 2, we see that from a practical viewpoint the estimator (22) is strong-minimax for a wider range of β. This is rather useful property, since the noise level σ is usually known only approximately.

Notice also that $\bar{\mu}_\beta(Z)$ in (21) is the minimax estimator of μ for any $\beta \in [0, 1/2]$, i.e.,

$$\inf_{\tilde{\mu}}\sup_{\mu>0}\mathbf{E}_\mu[\tilde{\mu}(Z) - \mu]^2 = \sup_{\mu>0}\mathbf{E}_\mu[\bar{\mu}_\beta(Z) - \mu]^2 = \sigma^2, \quad \beta \in [0, 1/2].$$

Along with strong-minimax estimates of $1/\mu$ we will need in the sequel strong-minimax estimates of $1/\mu^2$ defined as follows.

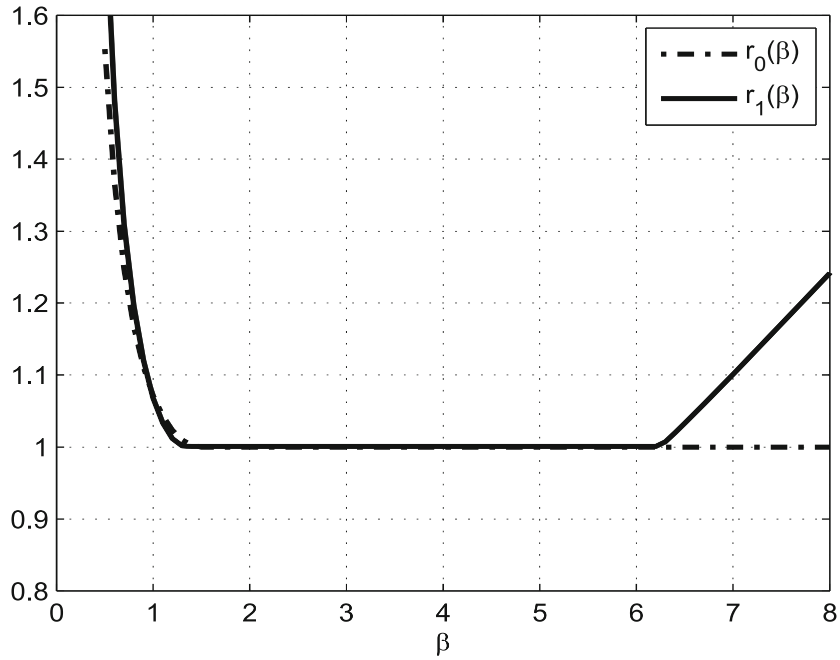

Fig. 2. Risk functions $r_0(\beta)$ and $r_1(\beta)$ for the estimator (22).

Definition 2. An estimator $\bar{\mu}^{-2}(z)$ of $1/\mu^2$ is called strong-minimax if

- $$\sup_{\mu>0} \mu^2 \mathbf{E}_\mu \big[\bar{\mu}^{-2}(z)\mu^2 - 1\big]^2 = 4\sigma^2; \tag{23}$$

- $$\sup_{\mu>0} \mathbf{E}_\mu [\bar{\mu}^{-2}(z)\mu^2]^2 = 1. \tag{24}$$

Recall that the usual minimax estimator $\bar{\mu}^{-2}(z)$ of $1/\mu^2$ and its minimax risk are defined by

$$r_2(\sigma) \stackrel{\text{def}}{=} \inf_{\tilde{\mu}^{-2}} \sup_{\mu>0} \mu^6 \mathbf{E}_\mu \left[\tilde{\mu}^{-2}(z) - \frac{1}{\mu^2}\right]^2 = \sup_{\mu>0} \mu^6 \mathbf{E}_\mu \left[\bar{\mu}^{-2}(z) - \frac{1}{\mu^2}\right]^2,$$

where the infimum is taken over all measurable functions $\tilde{\mu}^{-2}(\cdot)\colon \mathbb{R}^1 \to \mathbb{R}^+$.

The next lemma bounds from below the minimax risk $r_2(\sigma)$.

Lemma 4.

$$r_2(\sigma) \ge 4\sigma^2.$$

The proof of this lemma is quite similar to that of Lemma 1 and therefore it is omitted.

In order to show that the set of strongly-minimax estimates of $1/\mu^2$ is nonempty, we study numerically the family of the following estimates (see Eqs. (21) and (22) for a motivation):

$$\tilde{\mu}_\beta^{-2}(z) = \frac{1}{[\bar{\mu}_\beta(z)]^2} = \frac{1}{\beta^2\sigma^4}\left[\sqrt{\frac{z^2}{4} + \beta\sigma^2} - \frac{z}{2}\right]^2.$$

Lemma 5. *There exist constants $\tilde{\beta}_\circ$, $\tilde{\beta}^\circ$ such that $\tilde{\mu}_\beta^{-2}(z)$ is strong-minimax for any $\beta \in [\tilde{\beta}_\circ, \tilde{\beta}^\circ]$.*

The risk functions

$$r_0(\beta) = \sup_{\mu>0} \mathbf{E}_\mu [\mu^2 \tilde{\mu}_\beta^{-2}(z)]^2 \qquad \text{and} \qquad r_1(\beta) = \frac{1}{4} \sup_{\mu>0} \mu^2 \mathbf{E}_\mu [\mu^2 \tilde{\mu}_\beta^{-2}(z) - 1]^2$$

related to the estimate $\tilde{\mu}^{-2}(z)$ are plotted in Fig. 3. From this figure we see that $\tilde{\beta}^\circ \approx 2.5$ and $\tilde{\beta}_\circ \approx 8.8$.

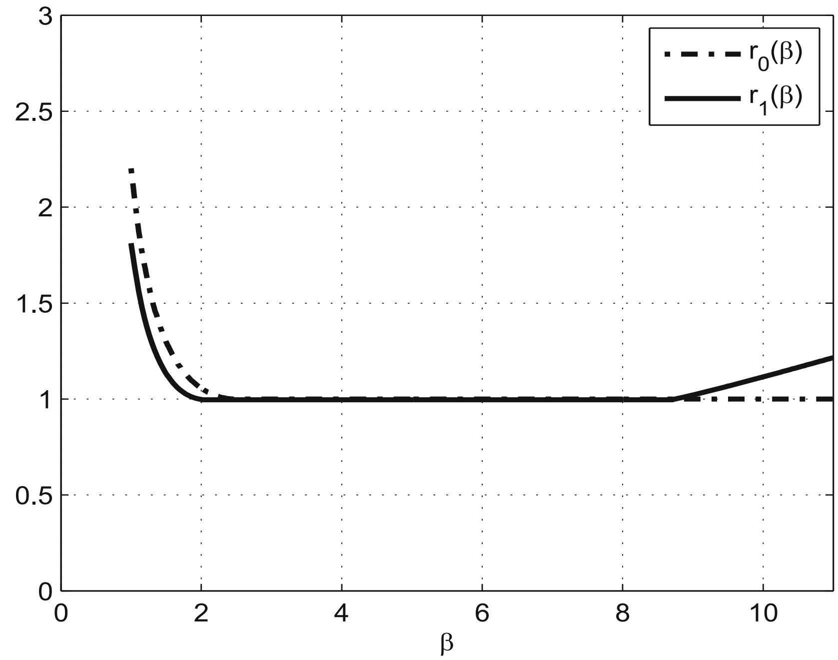

Fig. 3. Risk functions $r_0(\beta)$ and $r_1(\beta)$ for $\tilde{\mu}_\beta^{-2}(z)$.

2.2. Roughness Penalty Inversion

One of the most standard ways to construct good estimates of high-dimensional vectors θ in (1) is based on the roughness penalty approach. Suppose θ_k are independent zero mean Gaussian random variables with zero mean and

$$\mathbf{E}\theta_k^2 = \Sigma_k^2, \quad k = 1, \dots, p.$$

Let $\bar{b}^{-1}(z_k)$ be a strong-minimax estimate of $1/b_k$ (see (13) and (14)). Then we estimate unknown b_k by $1/\bar{b}^{-1}(z_k)$ and thus we estimate θ_k in Model (1) as follows:

$$\bar{\theta}_k(y_k, z_k) = \arg\min_\theta \biggl\{ -\frac{1}{2\epsilon^2}\biggl[\frac{\theta}{\bar{b}^{-1}(z_k)} - y_k\biggr]^2 - \frac{\theta^2}{2\Sigma_k^2}\biggr\}.$$

It can be easily seen that

$$\bar{\theta}_k(y_k, z_k) = \frac{\bar{b}^{-1}(z_k)}{1 + \epsilon^2\Sigma_k^{-2}[\bar{b}^{-1}(z_k)]^2} y_k.$$

In Model (2) we estimate θ based on the same idea, i.e.,

$$\tilde{\theta}_k(y_k', z_k) = \arg\min_\theta \biggl\{ -\frac{\tilde{b}^{-2}(z_k)}{2\epsilon^2}\biggl[\frac{\theta}{\tilde{b}^{-2}(z_k)} - y_k\biggr]^2 - \frac{\theta^2}{2\Sigma_k^2}\biggr\},$$

or, equivalently,

$$\tilde{\theta}_k(y_k, z_k) = \frac{\tilde{b}^{-2}(z_k)}{1 + \epsilon^2\Sigma_k^{-2}\tilde{b}^{-2}(z_k)} y_k'. \tag{25}$$

It is assumed that in the above equations $\tilde{b}^{-2}(z_k)$ is a strong-minimax estimate of $1/b_k^2$.

Our goal is to show that $\bar{\theta}(y, z)$ and $\tilde{\theta}(y, z)$ can mimic the pseudo-estimate in Models (1) and (2)

$$\hat{\theta}_k^\circ(y_k) = h_k^\circ \frac{y_k}{b_k}, \tag{26}$$

where

$$h_k^\circ = \frac{1}{1+\epsilon^2\Sigma_k^{-2}b_k^{-2}}.$$

Let us emphasize that $\hat{\theta}_k^\circ(y_k)$ is the roughness penalty estimate constructed assuming that b_k are known exactly.

Theorem 1. *Let $\bar{b}^{-1}(z_k)$ be a strong-minimax estimate of $1/b_k$. Then*

$$\big[\mathbf{E}\|\bar{\theta}(y,z)-\theta\|^2\big]^{1/2} \le \big[\mathbf{E}\|\hat{\theta}^\circ(y)-\theta\|^2\big]^{1/2} + \bigg\{\sum_{k=1}^p h_k^\circ\bigg[\sigma^2\frac{\theta_k^2}{b_k^2}+\frac{\epsilon^2}{b_k^2}\min\Big(1,\frac{\sigma^2}{b_k^2}\Big)\bigg]\bigg\}^{1/2}. \tag{27}$$

For the projection method $\bar{\theta}_k(y,z)=\mathbf{1}\{k\le W\}\bar{b}^{-1}(z_k)y_k$ the following inequality holds:

$$\big[\mathbf{E}\|\bar{\theta}(y,z)-\theta\|^2\big]^{1/2} \le \big[\mathbf{E}\|h^\circ\cdot y-\theta\|^2\big]^{1/2} + \sigma\bigg[\sum_{k=1}^W\frac{\theta_k^2}{b_k^2}\bigg]^{1/2}, \tag{28}$$

where $h_k^\circ=\mathbf{1}\{k\le W\}$.

Proof. Notice that $\bar{\theta}(y,z)$ admits the following decomposition

$$\bar{\theta}_k(y_k,z_k)=\frac{1}{b_k}\,\frac{b_k\bar{b}^{-1}(z_k)}{1+\epsilon^2\Sigma_k^{-2}b_k^{-2}\,[b_k\bar{b}^{-1}(z_k)]^2}\,y_k.$$

Denote for brevity

$$\rho_k=\epsilon^2\Sigma_k^{-2}b_k^{-2},\quad \zeta_k=b_k\bar{b}^{-1}(z_k),\quad h_k^\circ=\frac{1}{1+\rho_k},\quad \bar{h}_k=\frac{\zeta_k}{1+\rho_k\zeta_k^2}.$$

Let us begin with analyzing the projection method. In this case

$$\Sigma_k^2=\begin{cases}\infty, & k\le W,\\ 0, & k>W,\end{cases}$$

where W is a given projection frequency. So, we obviously obtain

$$h_k^\circ=\mathbf{1}\{k\le W\}\qquad\text{and}\qquad \bar{h}_k=\zeta_k\mathbf{1}\{k\le W\}.$$

Therefore it can be easily seen that

$$\mathbf{E}\|\theta-\hat{\theta}_k^\circ(y)\|^2=\sum_{k>W}\theta_k^2+\epsilon^2\sum_{k=1}^W\frac{1}{b_k^2},$$

and by the strong-minimax property of $\bar{b}_k^{-1}(z_k)$ (see (13) and (14)) we obtain

$$\begin{aligned}\mathbf{E}\|\theta-\bar{\theta}(y,z)\|^2 &= \sum_{k>W}\theta_k^2+\mathbf{E}\sum_{k=1}^W(1-\zeta_k)^2\theta_k^2+\epsilon^2\mathbf{E}\sum_{k=1}^W\frac{\zeta_k^2}{b_k^2}\\ &\le \sum_{k>W}\theta_k^2+\epsilon^2\sum_{k=1}^W\frac{1}{b_k^2}+\sigma^2\sum_{k=1}^W\frac{\theta_k^2}{b_k^2}=\mathbf{E}\|\theta-\hat{\theta}^\circ(y)\|^2+\sigma^2\sum_{k=1}^p h_k^{\circ 2}\frac{\theta_k^2}{b_k^2},\end{aligned}$$

thus proving (28).

In the general case, to control the risk of $\bar{\theta}(y,z)$, we make use of the following equation:

$$\mathbf{E}\|\theta-\bar{\theta}(y,z)\|^2=\mathbf{E}\|(1-\bar{h})\cdot\theta\|^2+\epsilon^2\sum_{k=1}^p b_k^{-2}\mathbf{E}\bar{h}_k^2.$$

We begin with upper-bounding the last term in this equation. With a simple algebra one obtains

$$\begin{aligned}
\mathbf{E}\bar{h}_k^2 &= h_k^2 \mathbf{E}\biggl[\frac{\zeta_k(1+\rho_k)}{1+\rho_k\zeta_k^2}\biggr]^2 \\
&\le h_k^2\mathbf{E}\zeta_k^2\mathbf{1}\{\zeta_k \ge 1\} + h_k^2\mathbf{E}\biggl[\zeta_k + (1-\zeta_k^2)\frac{\rho_k\zeta_k}{1+\rho_k\zeta_k^2}\biggr]^2\mathbf{1}\{\zeta_k < 1\} \\
&\le h_k^2\mathbf{E}\zeta_k^2\mathbf{1}\{\zeta_k \ge 1\} + h_k^2\mathbf{E}\biggl[\zeta_k + (1-\zeta_k)\frac{2\rho_k\zeta_k}{1+\rho_k\zeta_k^2}\biggr]^2\mathbf{1}\{\zeta_k < 1\} \\
&\le h_k^2\mathbf{E}\zeta_k^2\mathbf{1}\{\zeta_k \ge 1\} + h_k^2\mathbf{E}\bigl[\zeta_k + \sqrt{\rho_k}(1-\zeta_k)\bigr]^2\mathbf{1}\{\zeta_k < 1\}. \qquad (29)
\end{aligned}$$

In deriving the above inequality it was used that

$$\max_{x\ge 0}\frac{x}{1+\rho_k x^2} = \frac{1}{2\sqrt{\rho_k}}.$$

Next we continue (29) with the help of $\mathbf{E}(1-\zeta_k)_+^2 \le \min\{1, \sigma^2 b_k^{-2}\}$, which easily follows from the strong-minimax property of $\bar{b}_k^{-1}(z_k)$. Using

$$(x+y)^2 \le (1+z)x^2 + \biggl(1+\frac{1}{z}\biggr)y^2,\ z>0, \qquad (30)$$

we obtain for any $z > 0$

$$\begin{aligned}
\mathbf{E}\bar{h}_k^2 &\le (1+z)h_k^2 + \biggl(1+\frac{1}{z}\biggr)\rho_k h_k^2\min\{1,\sigma^2 b_k^{-2}\} \\
&\le (1+z)h_k^2 + \biggl(1+\frac{1}{z}\biggr)h_k\min\{1,\sigma^2 b_k^{-2}\}. \qquad (31)
\end{aligned}$$

Now, we proceed with upper-bounding $\mathbf{E}(1-\bar{h}_k)^2$. Obviously, we have

$$1-\tilde{h}_k = \frac{\rho_k}{1+\rho_k} + \frac{1}{1+\rho_k}(\zeta_k - 1)\frac{\rho_k\zeta_k - 1}{1+\rho_k\zeta_k^2}. \qquad (32)$$

Notice also that

$$\frac{|\rho_k\zeta_k - 1|}{1+\rho_k\zeta_k^2} = \rho_k\frac{|\rho_k\zeta_k - 1|}{\rho_k + (\rho_k\zeta_k)^2} \le \rho_k\max_{x\ge 0}\frac{|x-1|}{\rho_k + x^2}. \qquad (33)$$

One can also check with a simple algebra that

$$\max_{x\ge 0}\frac{x-1}{\rho_k+x^2} = \frac{1}{2+2\sqrt{1+\rho_k}},$$

and thus

$$\begin{aligned}
\rho_k\max_{x\ge 0}\frac{|x-1|}{\rho_k+x^2} &= \rho_k\max\biggl\{\frac{1}{\rho_k}, \frac{1}{2+2\sqrt{1+\rho_k}}\biggr\} \\
&= \max\biggl\{1, \frac{\sqrt{1+\rho_k}-\rho_k}{2}\biggr\} \le \sqrt{1+\rho_k}. \qquad (34)
\end{aligned}$$

Hence, combining (32)–(34) with (30) and with the strong-minimax property of $\bar{b}^{-1}(z_k)$, we arrive at the following inequality:

$$\mathbf{E}[1-\bar{h}_k]^2 \le (1+z)[1-h_k]^2 + \sigma^2\biggl(1+\frac{1}{z}\biggr)h_k b_k^{-2} \qquad (35)$$

that holds for any $z > 0$.

Thus with (35) and (31) we get

$$\mathbf{E}\|\theta-\bar{\theta}(y,z)\|^2 \le (1+z)\mathbf{E}\|\theta-h\cdot y\|^2 + \left(1+\frac{1}{z}\right)\sum_{k=1}^{p} h_k\left[\sigma^2\frac{\theta_k^2}{b_k^2}+\frac{\epsilon^2}{b_k^2}\min\left\{1,\frac{\sigma^2}{b_k^2}\right\}\right].$$

Finally, minimizing the right-hand side of this equation w.r.t. $z>0$, we finish the proof of (27). □

The next theorem controls the performance of the roughness penalty method in Model (2).

Theorem 2. *Let $\tilde{b}^{-2}(z_k)$ be a strong-minimax estimate of $1/b_k^2$. Then*

$$\left[\mathbf{E}\|\tilde{\theta}(y,z)-\theta\|^2\right]^{1/2} \le \left[\mathbf{E}\|\hat{\theta}^\circ(y)-\theta\|^2\right]^{1/2} + \left\{\sum_{k=1}^{p} h_k^{\circ 2}\left[4\sigma^2\frac{\theta_k^2}{b_k^2}+\frac{\epsilon^2}{b_k^2}\min\left(1,\frac{4\sigma^2}{b_k^2}\right)\right]\right\}^{1/2}. \tag{36}$$

For the projection estimate $\tilde{\theta}_k(y,z)=\mathbf{1}\{k\le W\}\tilde{b}^{-2}(z_k)y_k'$ the following inequality holds:

$$\left[\mathbf{E}\|\tilde{\theta}(y,z)-\theta\|^2\right]^{1/2} \le \left[\mathbf{E}\|h^\circ\cdot y-\theta\|^2\right]^{1/2} + 2\sigma\left[\sum_{k=1}^{W}\frac{\theta_k^2}{b_k^2}\right]^{1/2}, \tag{37}$$

where $h_k^\circ=\mathbf{1}\{k\le W\}$.

Proof. In view of (25), we can decompose $\tilde{\theta}(y',z)$ as follows:

$$\tilde{\theta}_k(y_k',z_k)=\frac{\tilde{b}^{-2}(z_k)b_k^2}{1+\epsilon^2\Sigma_k^{-2}\tilde{b}^{-2}(z_k)}\cdot\frac{y_k'}{b_k^2}.$$

Denote for brevity

$$\zeta_k=\tilde{b}^{-2}(z_k)b_k^2,\quad \rho_k=\frac{\epsilon^2}{\Sigma_k^2 b_k^2},\quad \tilde{h}_k=\frac{\zeta_k}{1+\rho_k\zeta_k}.$$

With this notation we have

$$\mathbf{E}[\theta_k-\tilde{\theta}_k(y_k',z_k)]^2=\mathbf{E}[1-\tilde{h}_k]^2\theta_k^2+\epsilon^2 b_k^{-2}\mathbf{E}\tilde{h}_k^2. \tag{38}$$

We begin with upper-bounding the right-hand side of this equation for the projection estimate with the projection frequency W. For this estimate,

$$\tilde{h}_k=\zeta_k\mathbf{1}\{k\le W\}.$$

Therefore by the strong-minimax property of $\tilde{b}^{-2}(z_k)$ we obtain from (38)

$$\mathbf{E}\|\theta-\tilde{\theta}(y',z)\|^2 \le \sum_{k=W+1}^{p}\theta_p^2+\epsilon^2\sum_{k=1}^{W}\frac{1}{b_k^2}+4\sigma^2\sum_{k=1}^{W}\frac{\theta_k^2}{b_k^2},$$

thus proving (37).

Let us now turn to the general case. We begin with controlling the bias term in the risk decomposition (38). Using (30) and the strong minimax property of $\tilde{b}^{-2}(z_k)$, we obtain

$$\begin{aligned}\mathbf{E}[1-\tilde{h}_k]^2 &= \mathbf{E}\left[\frac{\rho_k}{1+\rho_k}+\frac{1-\zeta_k}{(1+\rho_k\zeta_k)(1+\rho_k)}\right]^2\\ &\le (1+z)\left[\frac{\rho_k}{1+\rho_k}\right]^2+\left(1+\frac{1}{z}\right)\mathbf{E}\left[\frac{1-\zeta_k}{(1+\rho_k\zeta_k)(1+\rho_k)}\right]^2\\ &\le (1+z)[1-h_k^\circ]^2+4\sigma^2\left(1+\frac{1}{z}\right)\frac{h_k^{\circ 2}}{b_k^2}.\end{aligned} \tag{39}$$

With the same arguments we upper-bound the variance term

$$
\begin{aligned}
\mathbf{E}\tilde{h}_k^2 &= h_k^{\circ 2}\mathbf{E}\biggl[\frac{\zeta_k(1+\rho_k)}{1+\rho_k\zeta_k}\biggr]^2 \\
&\le h_k^{\circ 2}\mathbf{E}\zeta_k^2\mathbf{1}\{\zeta_k \ge 1\} + h_k^{\circ 2}\mathbf{E}\biggl[\zeta_k + (1-\zeta_k)\frac{\rho_k\zeta_k}{1+\rho_k\zeta_k}\biggr]^2\mathbf{1}\{\zeta_k < 1\} \\
&\le h_k^{\circ 2}\mathbf{E}\zeta_k^2\mathbf{1}\{\zeta_k \ge 1\} + h_k^{\circ 2}\mathbf{E}\bigl[\zeta_k + (1-\zeta_k)\bigr]^2\mathbf{1}\{\zeta_k < 1\} \\
&\le (1+z)h_k^{\circ 2} + \biggl(1+\frac{1}{z}\biggr)h_k^{\circ 2}\min\biggl\{1, \frac{4\sigma^2}{b_k^2}\biggr\}. \qquad (40)
\end{aligned}
$$

Finally, combining (38), (40), and (39), we finish the proof. □

2.3. Minimax Multivariate Inversion

Since the upper bounds in Theorems 1 and 2 are almost equivalent but Theorem 2 deals with a more general statistical model, we will focus in what follows on Model (2). With the help of Theorem 2 one can easily compute the maximal risk $\mathbf{E}\|\tilde{\theta}(y', z) - \theta\|^2$ over the ellipsoid

$$
\Theta = \biggl\{\theta\colon \sum_{k=1}^{p}\theta_k^2 a_k^2 \le 1\biggr\},
$$

where $\{a_k^2, k = 1, \dots, p\}$ is a given monotone sequence $a_1^2 \le a_2^2 \le \dots \le a_p^2$.

Theorem 3. *The maximal risk of $\tilde{\theta}(y', z)$ in (25) is upper-bounded as follows:*

$$
\biggl\{\sup_{\theta\in\Theta}\mathbf{E}\|\tilde{\theta}(y', z) - \theta\|^2\biggr\}^{1/2} \le \sqrt{R(\Sigma, \Theta)} + \sqrt{R^+(\Sigma, \Theta)},
$$

where

$$
\begin{aligned}
R(\Sigma, \Theta) &= \epsilon^4 \max_k \frac{1}{(\epsilon^2 + b_k^2\Sigma_k^2)^2 a_k^2} + \epsilon^2\sum_{k=1}^{p}\frac{\Sigma_k^4 b_k^2}{(\epsilon^2 + \Sigma_k^2 b_k^2)^2}, \\
R^+(\Sigma, \Theta) &= 4\sigma^2 \max_k \frac{\Sigma_k^4 b_k^2}{(\epsilon^2 + \Sigma_k^2 b_k^2)^2 a_k^2} + \epsilon^2\sum_{k=1}^{p}\frac{\Sigma_k^4 b_k^2}{(\epsilon^2 + \Sigma_k^2 b_k^2)^2}\min\biggl\{1, \frac{4\sigma^2}{b_k^2}\biggr\}.
\end{aligned}
$$

Proof. This follows immediately from (36) combined with

$$
h_k^\circ = \frac{b_k^2\Sigma_k^2}{\epsilon^2 + b_k^2\Sigma_k^2}, \qquad \sup_{\theta\in\Theta}\|(1-h^\circ)\cdot\theta\|^2 \le \max_k (1-h_k^\circ)^2 a_k^{-2},
$$

$$
\sup_{\theta\in\Theta}\sum_{k=1}^{p} h_k^{\circ 2}\frac{\theta_k^2}{b_k^2} \le \max_k \frac{h_k^{\circ 2}}{b_k^2 a_k^2}.
$$

□

The minimax risk of the projection method can be controlled with the help of the following theorem.

Theorem 4. *Let $\tilde{\theta}_{\mathrm{pr}}(y_k', z_k) = \mathbf{1}\{k \le W\}\tilde{b}^{-2}(z_k)y_k'$, then*

$$
\biggl\{\sup_{\theta\in\Theta}\mathbf{E}\|\tilde{\theta}_{\mathrm{pr}}(y', z) - \theta\|^2\biggr\}^{1/2} \le \sqrt{R_{\mathrm{pr}}(W, \Theta)} + \sqrt{R_{\mathrm{pr}}^+(W, \Theta)},
$$

where

$$
R_{\mathrm{pr}}(W, \Theta) = a_{W+1}^{-2} + \epsilon^2\sum_{k=1}^{W}\frac{1}{b_k^2}, \qquad R_{\mathrm{pr}}^+(W, \Theta) = 4\sigma^2 \max_{k\in[1,W]}\frac{1}{b_k^2 a_k^2}.
$$

Proof. This follows immediately from (37). □

Example. We illustrate the above theorem with a simple example, assuming that $p = \infty$ and

$$b_k^2 = B^2 k^{-2q}, \qquad a_k^2 = A^{-2} k^{2g}, \quad k = 1, 2, \ldots.$$

The computation of the risk of the spectral cut-off method in this case is very simple. We have

$$R_{\rm pr}(W, \Theta) = \frac{A^2}{(W+1)^{2q}} + \frac{\epsilon^2}{B^2} \sum_{k=1}^{W} k^{2g}, \qquad R_{\rm pr}^{+}(W, \Theta) = \frac{4\sigma^2 A^2 W^{2(g-q)_+}}{B^2},$$

where $(x)_+ = \max(0, x)$. Very often, we are interested in the minimax projection bandwidth minimizing $R_{\rm pr}(W, \Theta)$. This bandwidth can be easily computed for small ϵ, namely,

$$W^\circ = \arg\min_W R_{\rm pr}(W, \Theta) = (1 + o(1)) \left(\frac{2qA^2B^2}{\epsilon^2} \right)^{1/(1+2q+2g)}, \quad \epsilon \to 0,$$

and therefore as $\epsilon \to 0$

$$\min_W R_{\rm pr}(W, \Theta) = (1 + o(1)) \left(\frac{1}{2q+1} + \frac{1}{2q} \right) \frac{\epsilon^2}{B^2} \left(\frac{2qA^2B^2}{\epsilon^2} \right)^{(1+2q)/(1+2q+2g)}.$$

Notice also that

$$R_{\rm pr}^{+}(W^\circ, \Theta) = (1 + o(1)) \frac{4\sigma^2 A^2}{B^2} \left(\frac{2qA^2B^2}{\epsilon^2} \right)^{2(g-q)_+/(1+2q+2g)}, \quad \epsilon \to 0.$$

So, we see that when $q \geq g$ the excess risk $R_{\rm pr}^{+}(W^\circ, \Theta)$ has a parametric order σ^2.

This example shows, in particular, that one can construct good estimates of θ even in the case, where $\sigma^2 \gg \epsilon^2$. This prompts, for instance, that the upper bounds in Proposition 3.2 and Theorem 5.1 in [6] might be improved, since they are expressed in terms of $\max(\epsilon^2, \sigma^2)$.

Let us emphasize that the minimax projection bandwidth W° cannot be used in practice since it strongly depends on A^2 and q, which are hardly known. Therefore, in applications, only data-driven projection bandwidths can be used. Constructing good data-driven bandwidths is very important in applied statistics and we will provide a natural solution to this problem in a forthcoming paper.

Sometimes we are interested in computing Σ_k^2 resulting in asymptotically (as $\epsilon \to 0$) minimax estimators over Θ provided that b_k are assumed to be known. Recall that an asymptotically minimax estimate $\hat{\theta}_\epsilon(y)$ based on the observations

$$y_k = b_k \theta_k + \epsilon \xi_k, \quad k = 1, \ldots,$$

is defined by

$$\sup_{\theta \in \Theta} \mathbf{E} \|\hat{\theta}_\epsilon(y) - \theta\|^2 = (1 + o(1)) \inf_{\bar{\theta}} \sup_{\theta \in \Theta} \mathbf{E} \|\bar{\theta}(y) - \theta\|^2, \quad \epsilon \to 0,$$

where the infimum is taken over all estimates of θ. The theory of asymptotically minimax estimation over ellipsoids has been developed in the pioneering paper [9]. In particular, it follows from this paper that if

$$b_k^2 = (1 + o(1)) B^2 k^{2g}, \qquad a_k^2 = (1 + o(1)) A^{-2} k^{-2q} \quad \text{for } A, B, q, g \in (0, \infty),$$

as $k \to \infty$, then asymptotically minimax estimate of θ is given by (26) with

$$\Sigma_k^2 = \frac{\epsilon^2}{b_k^2} \left[\frac{|a_k|}{\mu} - 1 \right]_+, \quad \text{where } \mu \text{ is a root of } \quad \epsilon^2 \sum_{k=1}^{\infty} \frac{a_k^2}{b_k^2} \left[\frac{|a_k|}{\mu} - 1 \right]_+ = 1.$$

3. ACKNOWLEDGMENTS

This work was partially supported by RFBR research projects 13-01-12447 and 13-07-12111.

REFERENCES

1. T. Cai and P. Hall, "Prediction in Functional Linear Regression", Ann. Statist. **34**, 2159–2179 (2006).
2. L. Cavalier and N. W. Hengartner, "Adaptive Estimation for Inverse Problems with Noisy Operators", Inverse Problems **21**, 1345–1361 (2005).
3. X. Chen and M. Reiss, "On Rate Optimality for Ill-Posed Inverse Problems in Econometrics", Econometric Theory **27**, 497–521 (2011).
4. S. Efroimovich and V. Koltchinskii, "On Inverse Problems with Unknown Operators", IEEE Trans. Inform. Theory **47**, 2876–2894 (2001).
5. R. D. Gill and B. Y. Levit, "Applications of the Van Trees Inequality: A Bayesian Cramer–Rao Bound", Bernoulli **1**, 59–79 (1995).
6. M. Hoffmann and M. Reiss, "Nonlinear Estimation for Linear Inverse Problems with Error in the Operator", Ann. Statist. **36** (1), 310–336 (2008).
7. J. Johannes, S. van Bellegem, and A. Vanhems, "Convergence rates for ill-posed inverse problems with an unknown operator", Economic Theory **27**, 522–545 (2011).
8. C. Marteau, "Regularization of Inverse Problems with Unknown Operator", Math. Methods of Statist. **15**, 415–433 (2006).
9. M. S. Pinsker, "Optimal Filtering of Square Integrable Signals in Gaussian White Noise", Problems of Inform. Transmission **16**, 120–133 (1980).
10. H. L. Van Trees, *Detection, Estimation, and Modulation Theory* (Wiley, New York, 1968), Vol. 1.