A Linear Mini-Max Estimator for the Case of a Quartic Loss Function

by

K. Helmes[1)] *and C. Srinivasan*[2)]

1) Inst. for Operations Research, Humboldt University of Berlin; supported in part by NSF Grant DMS 94 04 990.
2) Dept. of Statistics, University of Kentucky.

Abstract

Let $Y(t)$ be a stochastic process on $[0,1]$ modeled as $dY_t = \theta(t)dt + dW(t)$, where $W(t)$ denotes a standard Wiener process, and $\theta(t)$ is an unknown function assumed to belong to a given set $\Theta \subset L^2[0,1]$. We consider the problem of estimating the value $L(\theta)$, where L is a known continuous linear function defined on Θ, using linear estimators of the form $\langle m, y\rangle = \int m(t)dY(t),\ m \in L^2[0,1]$. We solve the problem for the case of a *quartic* loss function, and compare the solution of the *quartic* case with the solution for the case of a *quadratic* loss function.

1 Introduction

We consider the following linear estimation problem. Let $Y(t)$ be a stochastic process on the interval [0,1] given by

$$dY(t) = \theta(t)dt + dW(t), \tag{1.1}$$

where W is a standard Wiener process and $\theta \in \Theta \subset L_2[0,1]$. A bounded continuous linear function L on Θ is specified, and the statistical problem is to estimate $L(\theta)$ by a linear estimator $\langle m, Y\rangle = \int_0^1 m(t)dY(t)$ for the case of a *quartic* loss function. The problem for a *quadratic* loss function was first addressed by Ibragimov and Hasminski [6], and they obtained a minimax theorem as well as a linear minimax estimator in this case; for related work cf. [2]–[5], [7], [12]–[13], and the literature cited therein. In [4] the authors use convex analysis to analyze the *quadratic* case. The basic idea of their proof for the quadratic case carries over to the case of a *quartic* loss function, although some of the arguments are slightly more involved. Following the ideas of Ibragimov and Hasminski we also derive a lower bound for the ratio of the minimax risks for all non-linear estimators versus all linear estimators for the quartic case. Since, by reparametrization, model (1.1) incorporates the case when the diffusion coefficient is different from 1 we shall restrict our analysis to the model class (1.1). Among the many applications of Gaussian white noise models we would like to mention non-linear regression problems and the estimation of a probability density; for details and further references cf. [11].

In Section 2 we present the needed results from convex analysis. Using convex analysis we give the proof of the minimax theorem for the finite dimensional case in Section 3. Section 4 contains infinite dimensional versions of this kind of minimax theorems which depend on the finite dimensional result. In Section 5 we derive the lower bound for the ratio of the minimax risks, and in Section 6 we analyze some examples.

2 Preliminaries and Notations

Let f be a real-valued convex function defined on a locally convex topological vector space X. The subdifferential of the function f at x_0 is defined to be the set

$$\partial f(x_0) = \{x' \in X' \colon \langle x - x_0, x' \rangle \le f(x) - f(x_0) \;\; \forall\, x \in X\}\,,$$

where X' denotes the dual of X. It is well-known that a convex function $f(x)$ attains a minimum at x^* if and only if $0 \in \partial f(x^*)$.

We will need the following results concerning the subdifferentials of convex functions. The first result is due to Moreau [10] and the second one, in its general form, is due to Ioffe and Levin [8].

Theorem 2.1 *Let f_1 and f_2 be two convex functions such that the relative interiors of their domains have a common point. Then*

$$\partial(f_1 + f_2)(x) = \partial f_1(x) + \partial f_2(x) \quad \textit{for all } x \in \textit{dom}\, f_1 \cap \textit{dom}\, f_2\,.$$

The next result is about the subdifferential of the pointwise supremum of convex functions. Let T be a compact space and $f_t(x)$ be a real function defined on $T \times X$ such that for each t, $f_t(x)$ is convex in x and for each x, continuous in t. Consider the convex function: $H(x) = \max_{t \in T} f_t(x)$. We are interested in the subdifferential of H at x_0. Towards this, let $T(x_0) = \{t \in T \colon\ f_t(x_0) = H(x_0)\}$, and define the set

$$W(x_0) = \{\pi \text{ is a probability on } T \text{ such that support } \pi \subset T(x_0)\}\,.$$

Theorem 2.2

$$\partial H(x_0) = \bigcup_{\pi \in W(x_0)} \left\{ \int_T \partial f_t(x_0) d\pi(t) \right\}$$

Proof: See Ioffe and Levin [8].

3 Finite Dimensional Minimax Theorem

In this section we prove the following minimax theorem for the finite dimensional case. Let $y = (y_1, \ldots, y_p)$ be a p-dimensional Gaussian random vector with mean $\theta \in \Theta \subset R^p$ and variance-covariance matrix $I_{p \times p}$. Consider estimating the quantity $\langle l, \theta \rangle = \sum_1^p l_i \theta_i$ where $l \in R^p$, by a linear estimator of the form $\langle m, y \rangle = \sum_i^p m_i y_i$ under the *quartic* loss $L(\theta, m) = \left(\sum l_i \theta_i - \sum m_i y_i \right)^4$. The risk of any such linear estimator is given by

$$R(\theta, m) = E\,[L(\theta, m)] = \langle \ell - m, \theta \rangle^4 + \binom{4}{2} \langle \ell - m, \theta \rangle^2 \|m\|^2 + 3(\|m\|^2)^2. \tag{3.1}$$

We shall show that when Θ is symmetric, compact, and convex, there is a linear estimator $\langle m^*, y \rangle$ minimax in the class of all linear estimators of the form $\langle m, y \rangle$, $m \in R^p$.

Theorem 3.1 *Suppose $\Theta \subset R^p$ is compact, convex and symmetric. Then*

$$\inf_{m \in R^p} \sup_{\theta \in \Theta} R(\theta, m) = \sup_{\theta \in \Theta} \inf_{m \in R^p} R(\theta, m)\,.$$

Proof: It suffices to prove that

$$\inf_{m} \sup_{\theta \in \Theta} R(\theta, m) \leq \sup_{\theta \in \theta} \inf_{m} R(\theta, m)\,. \tag{3.2}$$

Let us first analyze the r.h.s. of Ineq. (3.2). For the time being assume $\theta \in \Theta$ to be fixed. First, we solve the unrestricted minimization problem

$$\min_{m} R(\theta, m),$$

where $R(\theta, m)$ is given by Eq. (3.1). To this end, we exploit the necessary optimality condition $\frac{\partial R}{\partial m} = 0$. Since

$$\begin{aligned} \frac{\partial R}{\partial m} &= 4\langle \ell - m, \theta \rangle^3 (-\theta) + 6\|m\|^2 (2m) \\ &\quad + \binom{4}{2} \Big[2\langle \ell - m, \theta \rangle \|m\|^2 (-\theta) + 2\langle \ell - m, \theta \rangle^2 m \Big] \end{aligned} \tag{3.3}$$

the optimality condition yields

$$m_\theta = c_\theta z_\theta \theta, \tag{3.4}$$

where $c_\theta = \langle \ell - m_\theta, \theta \rangle$, and

$$z_\theta = \frac{\frac{c_\theta^2}{3} + \|m_\theta\|^2}{c_\theta^2 + \|m_\theta\|^2}\,. \tag{3.5}$$

In order not to overburden the notation we shall suppress the subscript θ whenever we are using the parameters c_θ and z_θ in the next couple of lines. Multiplying Eq. (3.4) by $\frac{1}{\|m\|}$, taking the norm on both sides of the equation, and using

$$\gamma = \gamma_\theta := \frac{|c|}{\|m\|}, \quad A := \frac{1}{\|\theta\|},$$

we get

$$1 = \gamma \left(\frac{\frac{\gamma^2}{3} + 1}{\gamma^2 + 1} \right) \|\theta\| \tag{3.6}$$

$$\iff \quad \gamma^3 - 3A\gamma^2 + 3\gamma - 3A = 0. \tag{3.7}$$

It follows from Eq. (3.7) that z, cf. Eq. (3.5), equals:

$$z = \frac{A}{\gamma},$$

and that z is the (unique) real root of

$$z^3 - z^2 + A^2 \left(z - \frac{1}{3} \right) = 0. \tag{3.8}$$

Since $c = \langle \ell - m, \theta \rangle$, and $m = cz\theta$ we have that, for fixed θ, $\min_m R(\theta, m)$ is attained at

$$m_\theta = \left\langle \ell, \frac{\theta}{\|\theta\|} \right\rangle \frac{z_\theta \|\theta\|^2}{1 + z_\theta \|\theta\|^2} \frac{\theta}{\|\theta\|} = \frac{z_\theta}{1 + z_\theta \|\theta\|^2} \langle \ell, \theta \rangle \theta \,. \tag{3.9}$$

Remark 1:

(i) Analyzing the polynomials $(z^3 - z)$ and $\big(- A^2(z - \frac{1}{3})\big)$ shows that Eq. (3.8) has a unique real root z_θ contained in the interval $(\frac{1}{3}, 1)$, and that z_θ is increasing in $\|\theta\|$.

(ii) The smaller the value of A (i.e. the larger the value of $\|\theta\|$), the closer z is to 1.

(iii) For $A \to \infty$ (i.e. $\|\theta\| \to 0$)), $z \to \frac{1}{3}$.

Substituting the expression on the r.h.s. of Eq. (3.9) in (3.1) we obtain, after some trivial modifications,

$$\sup_{\theta \in \Theta} \inf_m R(\theta, m) = \sup_{\theta \in \Theta} \left\{ 3 \left\langle \ell, \frac{\theta}{\|\theta\|} \right\rangle^4 \left(\frac{z_\theta \|\theta\|^2}{1 + z_\theta \|\theta\|^2} \right)^3 \left(\frac{1 + z_\theta^2 \|\theta\|^2}{z_\theta^2 \|\theta\|^2} \right) \right\} . \tag{3.10}$$

Now we consider the l.h.s. of Eq. (3.2). It is easily checked that $R(\theta, m)$, as given by Eq. (3.1), is a coercive, convex function in m. Consequently, $\inf_m \sup_{\theta \in \Theta} R(\theta, m)$ is attained. We denote a minimum point by m^*, i.e.

$$\inf_m \sup_{\theta\in\Theta} R(\theta,m) = \sup_{\theta\in\Theta} R(\theta,m^*)\,. \tag{3.11}$$

Since m^* is a minimum point, we must have

$$0 \in \partial\left[\sup_{\theta\in\Theta} R(\theta,m^*)\right]\,. \tag{3.12}$$

Condition (3.12) implies, exploiting Theorem 2.1, that

$$0 \in \partial\left[\sup_{\theta\in\Theta}\left\{\langle \ell - m^*,\theta\rangle^4 + \binom{4}{2}\langle \ell - m^*,\theta\rangle^2\|m^*\|^2\right\}\right] + 3\partial\left[(\|m^*\|^2)^2\right]\,.$$

By Theorem 2.2, and using the fact that the subdifferential of a Gateaux-differentiable function is a singleton set, we conclude:

$$0 \in \left\{\bigcup_{\pi\in W(m^*)}\int \partial\left[\langle \ell - m^*,\theta\rangle^4 + \binom{4}{2}\langle \ell - m^*,\theta\rangle^2\|m^*\|^2\right] d\pi(\theta) + 12\|m^*\|^2 m^*\right\}, \tag{3.13}$$

where

$$W(m^*) = \{\pi\colon\ \pi \text{ is a probability on } \Theta \text{ with support } \pi \subset T(m^*)\}$$

and

$$\begin{aligned} T(m^*) \;&=\; \left\{\theta\in\Theta \mid \langle \ell - m^*,\theta\rangle^4 + \binom{4}{2}\langle \ell - m^*,\theta\rangle^2\|m^*\|^2 \right. \\ &\qquad \left. = \sup_{\eta\in\Theta}\left\{\langle \ell - m^*,\eta\rangle^4 + \binom{4}{2}\langle \ell - m^*,\eta\rangle^2\|m^*\|^2\right\}\right\}\,. \end{aligned}$$

Note, $T(m^*)$ is non empty since Θ is compact. So there exists a probability measure $\pi_0 \in W(m^*)$ such that

$$0 = \int \frac{\partial R}{\partial m}\,(\theta,m^*)d\pi_0(\theta), \tag{3.14}$$

where $\frac{\partial R}{\partial m}\,(\theta,m^*)$ is given by Eq. (3.3), with m replaced by m^* on the r.h.s. of the equation. Next, for m^* and any $\theta\in T(m^*)$,

$$R(\theta,m^*) - 3\|m^*\|^4 = \langle \ell - m^*,\theta\rangle^2\left[\langle \ell - m^*,\theta\rangle^2 + 6\|m^*\|^2\right] \overset{\approx}{=}\text{const}$$

$$\Longrightarrow \ \langle \ell - m^*, \theta\rangle^2 = \widetilde{\text{const}}, \qquad \Longrightarrow \ |\langle \ell - m^*, \theta\rangle| = \text{const} =: c. \tag{3.15}$$

Let $s(\theta)$ denote the signum of $\langle \ell - m^*, \theta\rangle$. By Eq. (3.14) and the definition of c,

$$\begin{aligned} 12(c^2 + \|m^*\|^2)m^* &= 4(c^2 + 3\|m^*\|^2) \int |\langle \ell - m^*, \theta\rangle| s(\theta)\theta \, d\pi_0(\theta) \\ &= 4(c^2 + 3\|m^*\|^2)^2 \int s(\theta)\theta \, d\pi_0(\theta). \end{aligned} \tag{3.16}$$

Since Θ is a convex and symmetric (around zero) set,

$$\theta^* = \int s(\theta)\theta \, d\pi_0(\theta) \ \in \Theta. \tag{3.17}$$

Exploiting the linearity and continuity of $\langle \ell - m^*, \theta\rangle$, we obtain

$$\begin{aligned} c = \int c \, d\pi_0(\theta) &= \int |\langle \ell - m^*, \theta\rangle| d\pi_0(\theta) \\ &= \int \langle \ell - m^*, \theta\rangle s(\theta) d\pi_0(\theta) \\ &= \int \langle \ell - m^*, s(\theta)\theta\rangle d\pi_0(\theta) \\ &= \langle \ell - m^*, \theta^*\rangle . \end{aligned}$$

Hence, cf. Eqs. (3.4) and (3.5), we get

$$m^* = c^* z^* \theta^*, \tag{3.18}$$

where $c^* = c$, c defined by Eq. (3.15), and z^* is given by Eq. (3.5) with θ replaced by θ^*.

Repeating the arguments following Eq. (3.5), and with a slight abuse of notation, we finally get

$$R(\theta^*, m^*) = 3\left\langle \ell, \frac{\theta^*}{\|\theta^*\|}\right\rangle^4 \left(\frac{z^*\|\theta^*\|^2}{1 + z^*\|\theta^*\|^2}\right)^3 \left(\frac{1 + (z^*)^2\|\theta^*\|^2}{(z^*)^2\|\theta^*\|^2}\right) . \tag{3.19}$$

Therefore,

$$\begin{aligned} \inf_m \sup_{\theta\in\Theta} R(\theta, m) &= \sup_{\theta\in\Theta} R(\theta, m^*) \\ &= \sup_{\theta\in\Theta} \left\{ \langle \ell - m^*, \theta\rangle^4 \binom{4}{2} \langle \ell - m^*, \theta\rangle^2 \|m^*\|^2 + 3(\|m^*\|^2) \right\} \end{aligned}$$

$$
\begin{aligned}
&= R(\theta^*, m^*) \\
&= 3\left\langle \ell, \frac{\theta^*}{\|\theta^*\|}\right\rangle^4 \left(\frac{z^*\|\theta^*\|^2}{1+z^*\|\theta^*\|^2}\right)^3 \left(\frac{1+(z^*)^2\|\theta^*\|^2}{(z^*)^2\|\theta^*\|^2}\right) \\
&\leq \sup_{\theta\in\Theta}\left\{3\left\langle \ell, \frac{\theta}{\|\theta\|}\right\rangle^4 \left(\frac{z_\theta\|\theta\|^2}{1+z_\theta\|\theta\|^2}\right)^3 \left(\frac{1+z_\theta^2\|\theta\|^2}{z_\theta^2\|\theta\|^2}\right)\right\} \\
&= \sup_{\theta\in\Theta}\inf_{m} R(\theta, m)\,.
\end{aligned}
$$

$\diamond$

As an immediate consequence of the foregoing theorem and its proof we have the following representation result.

Theorem 3.2 *Let the supremum of the r.h.s. of Eq. (3.10) be attained at θ^*. Let z^* denote the (unique) real root of the polynomial equation (3.8) with $A = \frac{1}{\|\theta^*\|}$. The linear minimax estimator for the case of a quartic loss function is given by*

$$
m^* = \left\langle \ell, \frac{\theta^*}{\|\theta^*\|}\right\rangle \frac{z^*\|\theta^*\|^2}{1+z^*\|\theta^*\|^2}\,\frac{\theta^*}{\|\theta^*\|}\,. \tag{3.20}
$$

Example 1. Let Θ be the unit ball, i.e. $\Theta = \{\theta \in \Theta \mid \|\theta\| \leq 1\}$, and assume, without loss of generality, $\|\ell\| = 1$. It will be shown below that the supremum of

$$
\theta \longmapsto \left\langle \ell, \frac{\theta}{\|\theta\|}\right\rangle^4 \left(\frac{z_\theta\|\theta\|^2}{1+z_\theta\|\theta\|^2}\right)^3 \left(\frac{1+z_\theta^2\|\theta\|^2}{z_\theta^2}\right)\,, \tag{3.21}
$$

where z_θ is defined by Eq. (3.8), is attained on the boundary of Θ, more precisely at ℓ. Hence (numbers are printed to 4 decimals),

$$
z_\theta = 0.4425 \tag{3.22}
$$

and

$$
m^* = 0.3068\,\ell\,. \tag{3.23}
$$

The minimax risk is computed to be:

$$
R(\theta^*, m^*) = \left(\frac{z}{1+z}\right)^3 \left(\frac{1+z^2}{z^2}\right) = 0.5289\,. \tag{3.24}
$$

This result is consistent with the following analysis. Let m be fixed, and solve

$$
\sup_{\theta\in\Theta} R(\theta, m),
$$

with $R(\theta,m)$ given by Eq. (3.1). Since Θ is the unit ball the supremum is attained for

$$\theta = \frac{\ell - m}{\|\ell - m\|}.$$

Thus,

$$\sup_{\theta\in\Theta} R(\theta,m) = \|\ell - m\|^4 + 6\|\ell - m\|^2\|m\|^2 + 3\|m\|^4.$$

It is easy to see that the minimum (with respect to m) of the r.h.s. is attained for

$$m = \alpha\ell,$$

where α is the minimum of

$$q(\alpha) := (1-\alpha)^4 + 6(1-\alpha)^2\alpha^2 + 3\alpha^4.$$

The minimum is computed to be:

$$\frac{1}{\alpha} = 2 + \sqrt[3]{2} \;=\; 3.2599.$$

Hence,

$$\alpha = 0.3068,$$

which is consistent with formula (3.23).

It remains to be shown that the maximum of the function defined by Eq. (3.21) is attained for $\theta = \ell$. This follows once we have shown the function

$$F(\alpha) := \left(\frac{z(\alpha)\alpha^2}{1+z(\alpha)\alpha^2}\right)^3 \left(\frac{1+z^2(\alpha)\alpha^2}{z^2(\alpha)\alpha^2}\right) \tag{3.25}$$

to be monotone increasing in $\alpha \geq 0$; $z(\alpha)$ denotes the unique real root of Eq. (3.8), where $A = \frac{1}{\alpha}$.

To this end consider the function of two variables, $H(\alpha,z)$, $0 \leq \alpha < \infty$, $\frac{1}{3} < z \leq 1$,

$$H(\alpha,z) = \left(\frac{z\alpha^2}{1+z\alpha^2}\right)^3 \left(\frac{1+z^2\alpha^2}{z^2\alpha^2}\right). \tag{3.26}$$

By taking partial derivatives of H it is obvious that H is increasing in each variable (while the other variable is kept fixed). Hence, for $\alpha_1 \leq \alpha_2$, since $z(\alpha_1) \leq z(\alpha_2)$ (cf. Remark 1)

$$\begin{aligned} F(\alpha_1) \;&=\; \left(\frac{z(\alpha_1)\alpha_1^2}{1+z(\alpha_1)\alpha_1^2}\right)^3 \left(\frac{1+z^2(\alpha_1)\alpha_1^2}{z^2(\alpha_1)\alpha_1^2}\right) \\ &\leq\; \left(\frac{z(\alpha_2)\alpha_1^2}{1+z(\alpha_2)\alpha_1^2}\right)^3 \left(\frac{1+z^2(\alpha_2)\alpha_1^2}{z^2(\alpha_2)\alpha_1^2}\right) \end{aligned}$$

$$
\begin{aligned}
&\leq \left(\frac{z(\alpha_2)\alpha_2^2}{1+z(\alpha_2)\alpha_2^2}\right)^3 \left(\frac{1+z^2(\alpha_2)\alpha_2^2}{z^2(\alpha_2)\alpha_2^2}\right) \\
&= F(\alpha_2).
\end{aligned}
$$

We summarize the preceeding arguments in the following lemma.

Lemma 3.3 *The function*

$$
\theta \longmapsto \left\langle \ell, \frac{\theta}{\|\theta\|} \right\rangle^4 \left(\frac{z_\theta \|\theta\|^2}{1+z_\theta\|\theta\|^2}\right)^3 \left(\frac{1+z_\theta^2\|\theta\|^2}{z_\theta^2\|\theta\|^2}\right) ,
$$

attains its maximum on the boundary of the unit ball.

4 The Infinite Dimensional Case

We consider the infinite dimensional problem in this section. So in contrast to Section 2 $Y(t)$ is a stochastic process on the interval [0,1] given by the stochastic differential equation

$$
dY(t) = \theta(t)dt + dW(t), \quad 0 \leq t \leq 1 ,
$$

where $W(t)$ is a standard Wiener process, and the unknown parameter θ is assumed to lie in a set $\Theta \subset L_2[0,1]$. A continuous linear function L defined on Θ is given, and the problem is to estimate $L(\theta)$ by a linear estimator $\langle m, Y\rangle = \int_0^1 m(t)dY_t$ with the loss measured by the *quartic* function $(L(\theta) - \langle m, Y\rangle)^4$. The risk (i.e., expected loss) of any such estimator is given by

$$
\begin{aligned}
R(\theta, m) &= E\left[L(\theta, m)\right] = (L(\theta) - \langle m, \theta\rangle)^4 + \binom{4}{2}(L(\theta) - \langle m, \theta\rangle)^2\|m\|^2 + 3(\|m\|^2)^2, \\
&\qquad \theta \in \Theta, \quad m \in L_2, \qquad\qquad (4.1)
\end{aligned}
$$

where $\langle m, \theta\rangle = \int_0^1 m(t)d\theta(t)dt$ and $\|m\|$ is the L_2 norm of m.

We shall prove linear minimax theorems for estimating $L(\theta)$ under different sets of assumptions on Θ. *Throughout this section we assume that* Θ *is a symmetric (i.e., balanced) convex set.*

Theorem 4.1 *Let* Θ *be closed and bounded in* L_2. *Suppose* L *is weakly continuous on* Θ. *For each* $\theta \in \Theta$ *let* z_θ *denote the (unique) real root of the Eq.* (3.8)*, and let* c_θ *be defined by Eq.* (3.15)*. Then*

$$\begin{aligned}\inf_{m\in L_2}\sup_{\theta\in\Theta} R(\theta,m) &= \sup_{\theta\in\Theta}\inf_{m\in L_2} R(\theta,m)\\ &= \sup_{\theta\in\Theta}\left\{3\left(L\left(\frac{\theta}{\|\theta\|}\right)\right)^4\left(\frac{z_\theta\|\theta\|^2}{1+z_\theta\|\theta\|^2}\right)^3\left(\frac{1+z_\theta^2\|\theta\|^2}{z_\theta^2\|\theta\|^2}\right)\right\}. \quad (4.2)\end{aligned}$$

Moreover, if $\sup_{\theta\in\Theta}\left\{3\left(L\left(\frac{\theta}{\|\theta\|}\right)\right)^4\left(\frac{z_\theta\|\theta\|^2}{1+z_\theta\|\theta\|^2}\right)^3\left(\frac{1+z_\theta^2\|\theta\|^2}{z_\theta^2\|\theta\|^2}\right)\right\}$ *attains its supremum at* θ^*, *then the linear minimax estimator is given by*

$$\langle m^*, Y\rangle = \int_0^1 m^*(t)dY(t)$$

where

$$m^* = L\left(\frac{\theta^*}{\|\theta^*\|}\right)\frac{z^*\|\theta^*\|^2}{1+z^*\|\theta^*\|^2}\frac{\theta^*}{\|\theta^*\|}\,. \quad (4.3)$$

Proof: Observe that $R(\theta,m)$ is weakly continuous in θ and L_2 continuous in m. Moreover, Θ is weakly compact because it is bounded and closed in the L_2 topology. Hence, Ioffe and Levin's theorem (Theorem 2.2) is applicable to $\sup_\theta R(\theta,m)$ and the rest of the proof proceeds just as in Theorem 3.1 for the finite dimensional case.

The next theorem is, in a sense, a variant of the previous result. We replace the weak continuity assumption by a different condition.

Theorem 4.2 *Let* Θ *be closed with respect to the* L_2 *topology. Assume that* $L(\theta)$ *is bounded and continuous on* Θ. *Then,*

$$\begin{aligned}\inf_{m}\sup_{\theta\in\Theta} R(\theta,m) &= \sup_{\theta\in\Theta}\inf_{m} R(\theta,m)\\ &= \sup_{\theta\in\Theta}\left\{3\left(L\left(\frac{\theta}{\|\theta\|}\right)\right)^4\left(\frac{z_\theta\|\theta\|^2}{1+z_\theta\|\theta\|^2}\right)^3\left(\frac{1+z_\theta^2\|\theta\|^2}{z_\theta^2\|\theta\|^2}\right)\right\}. \quad (4.4)\end{aligned}$$

Proof: Since L is a continuous linear function on Θ, it has a continuous linear extension to the subspace spanned by Θ. This extension can be further extended to the entire space L_2 by the Hahn-Banach theorem. Let $\bar{L}$ denote this extension to L_2. Then, by virtue of the Riesz representation theorem, $\bar{L}$ has a representation of the form $\bar{L}(\theta) = \langle l,\theta\rangle = \int_0^1 l(t)\theta(t)dt$ such that $L(\theta) = \bar{L}(\theta) = \langle l,\theta\rangle$ on Θ. Thus, $L(\theta)$ is weakly continuous. Clearly, if Θ were bounded, the result would follow from the previous theorem (Theorem 4.2). In general, we will take advantage of this result to complete the proof. Let B_k be the ball of radius k in L_2. Then $\Theta\cap B_k$ is closed and bounded, and therefore, we have

$$\begin{aligned}\inf_m \sup_{\Theta\cap B_k} R(\theta,m) &= \sup_{\Theta\cap B_k} \inf_m R(\theta,m) \\ &\leq \sup_\theta \inf_m R(\theta,m) = \Delta\,. \end{aligned} \tag{4.5}$$

(**Note**: $\Delta = \sup_{\theta\in\Theta} \left\{ 3 \left\langle \ell, \frac{\theta}{||\theta||} \right\rangle^4 \left(\frac{z_\theta ||\theta||^2}{1+z_\theta||\theta||^2}\right)^3 \left(\frac{1+z_\theta^2||\theta||^2}{z_\theta^2||\theta||^2}\right) \right\}$)

Now, suppose

$$\inf_m \sup_\theta R(\theta,m) > \Delta + \delta \tag{4.6}$$

for some $\delta > 0$. Let, for every $k \geq 1$, $m_k \in L_2$ be such that

$$\sup_{\Theta\cap B_k} R(\theta,m_k) \leq \Delta + \frac{\delta}{10}\,. \tag{4.7}$$

Clearly, the m_k's are all contained in a bounded set, and therefore, have a subsequence converging weakly to, say, m^*. It follows from (4.6) that there exists a θ^* (depending on m^*) such that

$$E\left[L(\theta^*,m^*)\right] = \langle \ell - m^*, \theta^*\rangle^4 + \binom{4}{2} \langle \ell - m^*, \theta^* \rangle^2 \|m^*\|^2 + 3(\|m^*\|^2)^2 > \Delta + \frac{\delta}{5}. \tag{4.8}$$

Now, $\|m\|^2$ is a lower semi-continuous function with respect to weak convergence. Therefore, one can find a sufficiently large k, say κ, so that

$$R(\theta^*,m_\kappa) = E\left[L(\theta^*,m_k)\right] = \langle \ell - m_k, \theta^*\rangle^4 + \binom{4}{2} \langle \ell - m_k, \theta^*\rangle^2 \|m_k\|^2 + 3(\|m_k\|^2)^2 \geq \Delta + \frac{\delta}{5}\,,$$

and

$$\theta^* \in \Theta \cap B_\kappa\,,$$

which contradicts (4.8). This completes the proof of the theorem. $\diamond$

Taking up an idea of Ibragimov and Hasminski (1984), the preceding results can be generalized to the situation when Θ is not closed. The following theorem, which may be considered the main theorem of this section, treats this general case.

Theorem 4.3 *Suppose Θ is bounded and its finite dimensional sections are closed. Then*

$$\begin{aligned}\inf_m \sup_{\theta\in\Theta} R(\theta,m) &= \sup_{\theta\in\Theta} \inf_m R(\theta,m) \\ &= \sup_{\theta\in\Theta} \left\{ 3 \left(L\left(\frac{\theta}{\|\theta\|}\right)\right)^4 \left(\frac{z_\theta\|\theta\|^2}{1+z_\theta\|\theta\|^2}\right)^3 \left(\frac{1+z_\theta^2\|\theta\|^2}{z_\theta^2\|\theta\|^2}\right) \right\}. \end{aligned} \tag{4.9}$$

Proof: In view of the fact that $R(\theta, m)$ is convex in m and tends to infinity as $\|m\| \to \infty$, one can find a large enough positive constant $C > 0$ such that

$$\inf_{\|m\|\le C} \sup_{\theta\in\Theta} R(\theta, m) = \inf_{m} \sup_{\theta\in\Theta} R(\theta, m)$$

and

$$\begin{aligned}
\sup_{\theta\in\Theta} \inf_{\|m\|\le C} R(\theta, m) &= \sup_{\theta\in\Theta} \inf_{m} R(\theta, m) \\
&= \sup_{\theta\in\Theta} \left\{ 3\left(L\left(\frac{\theta}{\|\theta\|}\right)\right)^4 \left(\frac{z_\theta\|\theta\|^2}{1+z_\theta\|\theta\|^2}\right)^3 \left(\frac{1+z_\theta^2\|\theta\|^2}{z_\theta^2\|\theta\|^2}\right)\right\} \quad (4.10)
\end{aligned}$$

Therefore, setting $\mathcal{M} = \{m\colon \|m\| \le C\}$, it suffices to prove

$$\inf_{m\in\mathcal{M}} \sup_{\theta\in\Theta} R(\theta, m) = \sup_{\theta\in\Theta} \inf_{m\in\mathcal{M}} R(\theta, m)\,. \tag{4.11}$$

Now, for the rest of the proof, let H denote a finite dimensional subspace of $L_2[0,1]$ and $\mathcal{H}$ be the class of all finite dimensional subspaces H, i.e.,

$$\mathcal{H} = \{H \subset L_2[0,1]\colon H \text{ is a finite dimensional space}\}\ .$$

Set $\Theta_H = \Theta \cap H$. Clearly, $\Theta = \bigcup_{H\in\mathcal{H}} \Theta_H$ and each Θ_H is, by hypothesis, a closed, bounded, symmetric and convex finite dimensional set. Hence, by Theorem 3.1, we have for every $H \in \mathcal{H}$,

$$\inf_{m\in\mathcal{M}} \sup_{\theta\in\Theta_H} R(\theta, m) = \sup_{\theta\in\Theta_H} \inf_{m\in\mathcal{M}} R(\theta, m)\,. \tag{4.12}$$

Consequently,

$$\begin{aligned}
\sup_{\theta\in\Theta} \inf_{m\in\mathcal{M}} R(\theta, m) &= \sup_{H\in\mathcal{H}} \sup_{\theta\in\Theta_H} \inf_{m\in\mathcal{M}} R(\theta, m) \\
&= \sup_{H\in\mathcal{H}} \inf_{m\in\mathcal{M}} \sup_{\theta\in\Theta_H} R(\theta, m)\,. \quad (4.13)
\end{aligned}$$

Therefore, to establish (4.11), it is enough to show

$$\inf_{m\in\mathcal{M}} \sup_{H\in\mathcal{H}} \left[\sup_{\theta\in\Theta_H} R(\theta, m)\right] = \sup_{H\in\mathcal{H}} \inf_{m\in\mathcal{M}} \left[\sup_{\theta\in\Theta_H} R(\theta, m)\right] \tag{4.14}$$

because

$$\inf_{m\in\mathcal{M}} \sup_{\theta\in\Theta} R(\theta, m) = \inf_{m\in\mathcal{M}} \sup_{H\in\mathcal{H}} \left[\sup_{\theta\in\Theta_H} R(\theta, m)\right]\ .$$

Setting $G(m, H) = \sup_{\theta\in\Theta_H} R(\theta, m)$, (4.14) can be rewritten as

$$\inf_{m\in\mathcal{M}} \sup_{H\in\mathcal{H}} G(m, H) = \sup_{H\in\mathcal{H}} \inf_{m\in\mathcal{M}} G(m, H)\,. \tag{4.15}$$

We will establish (4.15) by appealing to a general minimax theorem.

Towards this end, equip $\mathcal{M}$ with the weak topology. Note that the weak topology coincides with weak* topology. Hence $\mathcal{M}$ is a compact convex Hausdorff space. Moreover, since $\|\cdot\|$ is lower semi-continuous with respect to the weak topology, $G(m,H)$ is a convex lower semicontinuous function in m for each $H \in \mathcal{H}$. Therefore, the conditions for the minimax theorem (Corollary 2 of Alpern and Gal [1]) are met and we have, F is defined below,

$$\inf_{m\in\mathcal{M}} \sup_{\pi\in\mathcal{H}^*} F(m,\pi) = \sup_{\pi\in\mathcal{H}^*} \inf_{m} F(m,\pi) \tag{4.16}$$

where $\mathcal{H}^*$ is the set of all discrete probability measures π on $\mathcal{H}$ with finite support and, if π_H is the probability π assigns to H,

$$F(m,\pi) = \sum_{H\in\operatorname{Supp}\pi} \pi_H\, G(m,H)\,.$$

Now, observe that for any $\pi \in \mathcal{H}^*$ there exists $H_\pi \in \mathcal{H}$ (e.g., $H_\pi =$ span$\{H\colon\ H$ in Supp $\pi\}$) such that

$$F(m,\pi) = \sum \pi_H\, G(m,H) \le G(m,H_\pi)\,, \tag{4.17}$$

and therefore,

$$\begin{aligned} \inf_{m\in\mathcal{M}} \sup_{\pi\in\mathcal{H}^*} F(m,\pi) &= \inf_{m\in\mathcal{M}} \sup_{H\in\mathcal{H}} G(m,H) \\ \sup_{\pi\in\mathcal{H}^*} \inf_{m\in\mathcal{M}} F(m,\pi) &\le \sup_{H\in\mathcal{H}} \inf_{m\in\mathcal{M}} G(m,H)\,. \end{aligned} \tag{4.18}$$

Finally, combining (4.16), (4.17) and (4.18), we have

$$\begin{aligned} \inf_{m\in\mathcal{M}} \sup_{H\in\mathcal{H}} G(m,H) = \inf_{m\in\mathcal{M}} \sup_{\pi\in\mathcal{H}^*} F(m,\pi) &= \sup_{\pi\in H^*} \inf_{m\in\mathcal{M}} F(m,\pi) \\ &\le \sup_{H\in\mathcal{H}} \inf_{m\in\mathcal{M}} G(m,H)\,, \end{aligned}$$

thus establishing (4.15). This completes the proof. $\diamond$

5 Linear versus Non-Linear Estimators

Naturally, the minimax risk when using all non-linear estimators is smaller than the risk for only linear estimators. Here we quantify for the case of a quartic loss function by how much. The results are variations of those obtained in [6] for the quadratic case.

Let

$$H(a) = \inf_{T} \sup_{|\theta|\le a} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} |T(x)-\theta|^4\, e^{-\frac{1}{2}(x-\theta)^2} dx, \tag{5.1}$$

where the infimum is taken over all measurable functions $T : I\!R \to I\!R$, and

$$\tilde{\Psi}(a,\sigma) = \inf_T \sup_{|\theta|\leq a} \frac{1}{\sigma}\frac{1}{\sqrt{2\pi}} \int_{I\!R} |T(x)-\theta|^4 \, e^{-\frac{1}{2}\frac{(x-\theta)^2}{\sigma^2}} dx. \tag{5.2}$$

Obviously, $H(a)$ is the value of the minimax quartic risk in the 1-dimensional problem of estimating the parameer $\theta \in [-a,a]$ from observing $\theta+\xi$, $\xi \sim \mathcal{N}(0,1)$.

The following identities are based on simple transformations of variables.

Lemma 5.1

$$\tilde{\Psi}(a,\sigma) = a^4\tilde{\Psi}\left(1,\frac{\sigma}{a}\right) = \sigma^4\tilde{\Psi}\left(\frac{a}{\sigma},1\right) = \sigma^4 H\left(\frac{a}{\sigma}\right).$$

Lemma 5.2 *The function $H(a)$ has the following properties:*

(i) $H(a)$ *is monotone increasing*

(ii) $H(\infty) = 3$

(iii) $H(a) \sim a^4$ *as* $a \to 0$.

Proof. Property (i) is obvious and equality (ii) is well known; see, for instance, the book by Ferguson on "Mathematical Statistics". To show (iii) we can mimic the proof given by Ibragimov and Hasminski of the similar kind of result for the quadratic case. Choosing $T \equiv 0$ we see that $H(a) \leq a^4$ for all $a \geq 0$. For sufficiently small a (see below), and $|T| \leq a$ we get:

$$\begin{aligned}
\sup_{|\theta|\leq a} E_\theta\Big[|\theta - T|^4\Big] &\geq \frac{1}{2}E_a\Big[|a-T|^4\Big] + \frac{1}{2}E_{-a}\Big[|-a-T|^4\Big] \\
&\geq \frac{1}{\sqrt{2\pi}}\int_{I\!R}\left(\frac{1}{2}\Big[T(x)-a\Big]^4 + \frac{1}{2}\Big[T(x)+a\Big]^4\right) e^{-\frac{x^2}{2}} dx \\
&\quad -14a^4\frac{1}{\sqrt{2\pi}}\int_{I\!R}\left|e^{-\frac{x^2}{2}} - e^{-\frac{(x-a)^2}{2}}\right| dx.
\end{aligned}$$

Since, in calculating $H(a)$ it suffies to consider only estimates $T(x)$ for which $|T(x)| \leq a$, and, trivially, $(T-a)^4 + (T+a)^4 \geq 2a^4$, we obtain

$$\begin{aligned}
\sup_{|\theta|\leq a} E_\theta\Big[\theta - T^2\Big] &\geq a^4 - 14a^4 e^{\frac{a^2}{2}}\left[\frac{3}{2}a^2 + \frac{2}{\sqrt{2\pi}}a\right] \\
&\geq a^4\Big(1 - O(a)\Big)
\end{aligned}$$

For the quartic case we define

$$\kappa := \inf_{||\theta||>0} \frac{H(||\theta||)}{3} \left(\frac{1+z_\theta||\theta||^2}{z_\theta||\theta||^2}\right)^3 \cdot \left(\frac{z_\theta^2||\theta||^2}{1+z_\theta^2||\theta||^2}\right), \tag{5.3}$$

where z_θ is defined by Eq. (3.8). It follows from property (iii) of Lemma 5.2 and from $\frac{1}{3} \le z_\theta \le 1$ (cf. Remark 1 in Section 3) that $\kappa > 0$. As to $\kappa \le 1$, note that

$$\kappa \le \lim_{||\theta||\to\infty} \frac{H(||\theta||)}{3} \left(\frac{1+z_\theta||\theta||^2}{z_\theta||\theta||^2}\right)^3 \cdot \left(\frac{z_\theta^2||\theta||^2}{1+z_\theta^2||\theta||^2}\right).$$

Thus, again by Lemma 5.2 and Remark 1, the assertion follows.

Theorem 5.3 *Let Θ be a closed convex subset of $L_2\big([0,1]\big)$. Then*

$$\begin{aligned} \sup_{\theta\in\Theta} \inf_m R(\theta,m) &\ge \inf_T \sup_{\theta\in\Theta} R(\theta,T) \\ &\ge \kappa \cdot \sup_{\theta\in\Theta} \left\{ 3\left\langle \ell, \frac{\theta}{||\theta||}\right\rangle^4 \left(\frac{z_\theta||\theta||^2}{1+z_\theta||\theta||^2}\right)^3 \left(\frac{1+z_\theta^2||\theta||^2}{z_\theta^2||\theta||^2}\right)\right\}. \end{aligned}$$

Proof. The proof of Theorem 5.3 is very much the same as the proof of Theorem 3 in [6], with the main difference being in the last step.

Let $\hat{L}$ by any estimate for $L(s)$ given observations of the form (1.1). Let us fix $s_0 \in \Theta$. Along with s_0 the functions θs_0, $|\theta| \le 1$, also belong to Θ. Let $dX_t = \theta s_0(t)dt + dW_t$, $X(0) = 0$, and consider the problem of estimating the 1-dimensional parameter θ given the observation

$$\begin{aligned} Y = \int_0^1 s_0(t)dX_t &= \theta \int_0^1 s_0^2(t)dt + \int_0^1 s_0(t)dW_t \\ &= \theta||s_0||^2 + ||s_0||\xi, \quad \xi \sim \mathcal{N}(0,1). \end{aligned}$$

By the properties of the functions $\tilde{\Psi}$ and H,

$$\inf_T \sup_{|\theta|\le 1} E_\theta\Big[|T-\theta|^4\Big] = \tilde{\Psi}\left(1, \frac{1}{||s_0||}\right) = \frac{1}{||s_0||^4} H\big(||s_0||\big). \tag{5.4}$$

Since Y is a sufficient statistic for the family of measures $\big(P_{\theta s_0}\big)_{|\theta|\le 1}$, see [6],

$$E_\theta\Big[\hat{L} \mid Y\Big] = E_0\Big[\hat{L} \mid Y\Big]. \tag{5.5}$$

Now, using the statistic $T_0[Y] := \frac{1}{L(s_0)} E_0\Big[\hat{L} \mid Y\Big]$ as an estimator for θ, and taking (5.4) and (5.5) into account

$$\frac{H(||s_0||)}{||s_0||^4} \le \frac{1}{L(s_0)^4} E_\theta\left[\Big(L(\theta s_0) - \hat{L}\Big)^4\right]. \tag{5.6}$$

Next, setting here $\theta = 1$, replacing s_0 by s and multiplying and dividing the left hand side of (5.6) by the mini-max risk

$$3\left\langle \ell, \frac{s}{||s||}\right\rangle^4 \left(\frac{z_s||s||^2}{1+z_s||s||^2}\right)^3 \left(\frac{1+z_s^2||s||^2}{z_s^2||s||^2}\right)$$

the theorem is proved.

In the following section we shall analyze some simple but useful examples. At the same time we illustrate the theorems and we show the kind of calculations needed in applications.

6 Examples

The first example to be discussed in this Section is a generalization of Example 1 (cf. Section 3).

Example 2. Let Θ be an ellipsoid, i.e. $\Theta = \{\theta \in I\!R^p \mid \theta'\Lambda\theta \leq 1\}$, where Λ is a symmetric, positive definite matrix. Because of Theorem 3.2 we may assume, without loss of generality, Λ to be diagonalized, i.e. $\Lambda = \mathrm{diag}\,(\lambda_1, \ldots, \lambda_p)$, $0 < \lambda_i$, $1 \leq i \leq p$, and $\|\ell\| = 1$. The linear minimax estimator for the case of a *quartic* loss function is given by any one of the following two equivalent expressions, (6.1) or (6.3):

$$m^* = \left(\mathrm{id} + \frac{1}{\zeta^*}\Lambda\right)^{-1} \ell\ , \tag{6.1}$$

where ζ^* is the (unique) solution in the interval $\left(\frac{1}{3}, 1\right)$ of the equation

$$\left(\zeta^3 - \zeta^2\right)\sum_{i=1}^{p} \frac{\ell_i^2}{(\zeta+\lambda_i)^2} + \left(\zeta - \frac{1}{3}\right)\sum_{i=1}^{p} \frac{\lambda_i \ell_i^2}{(\zeta+\lambda_i)^2} = 0\ , \tag{6.2}$$

or, equivalently,

$$m^* = \left\langle \ell, \frac{\theta^*}{\|\theta^*\|}\right\rangle \frac{\zeta^*\|\theta^*\|^2}{1+\zeta^*\|\theta\|^2}\,\frac{\theta^*}{\|\theta^*\|}\ , \tag{6.3}$$

where θ^* denotes the supremum of the r.h.s. of Eq. (3.10), and ζ^* is given by one of the equivalent characterizations of ζ^*, viz. Eq. (6.2), or Eq. (3.8) with $A = \frac{1}{\|\theta^*\|}$.

As we did in the case of Example 1, i.e. when $\Lambda = \mathrm{id}$, we shall give two different derivations of the linear minimax estimator for ellipsoidal constraints.

This time let us begin by analyzing, for fixed m, the problem

$$\sup_{\theta'\Lambda\theta\leq 1} R(\theta, m), \tag{6.4}$$

where $R(\theta, m)$ is again given by Eq. (3.1). Exploiting the necessary optimality conditions for this constrained optimization problem we obtain

$$\theta = \frac{\Lambda^{-1}(\ell - m)}{\sqrt{(\ell - m)'\Lambda^{-1}(\ell - m)}} \; . \tag{6.5}$$

Using this expression for θ when solving

$$\inf_{m} \sup_{\theta'\Lambda\theta\leq 1} R(\theta, m), \tag{6.6}$$

the minimization problem turns into

$$\inf_{m} \left\{ \left[(\ell - m)'\Lambda^{-1}(\ell - m)\right]^2 + 6\|m\|^2(\ell - m)'\Lambda^{-1}(\ell - m) + 3\|m\|^4 \right\} \; . \tag{6.7}$$

The necessary optimality conditions for this unconstrained optimization problem yield

$$m = \left(\mathrm{id} + \frac{1}{\zeta}\Lambda\right)^{-1} \ell \; , \tag{6.8}$$

and

$$\zeta = \frac{\|m\|^2 + \frac{1}{3}(\ell - m)'\Lambda^{-1}(\ell - m)}{\|m\|^2 + (\ell - m)'\Lambda^{-1}(\ell - m)} \; . \tag{6.9}$$

Multiplying Eq. (6.8) by $\left(\mathrm{id} + \frac{1}{\zeta}\Lambda\right)$ we obtain

$$\ell - m = \frac{1}{\zeta}\Lambda m \; . \tag{6.10}$$

Substituting $(\ell - m)$ from Eq. (6.10) for $(\ell - m)$ in Eqs. (6.5) and (6.9), and rearranging terms gives us

$$\theta = \frac{m}{\sqrt{m'\Lambda m}} \; , \tag{6.11}$$

and

$$\zeta = \frac{\zeta^2 + \frac{1}{3}\,\frac{m'\Lambda m}{\|m\|^2}}{\zeta^2 + \frac{m'\Lambda m}{\|m\|^2}} \; . \tag{6.12}$$

It follows from Eq. (6.11) that

$$B^2 := \frac{1}{\|\theta\|^2} = \frac{m'\Lambda m}{\|m\|^2} \; ; \tag{6.13}$$

Eq. (6.12) can thus be rewritten as

$$\zeta^3 - \zeta^2 + B^2\left(\zeta - \frac{1}{3}\right) = 0 \; . \tag{6.14}$$

Note, $B = B(\zeta)$ is a function of ζ. Thus, using the fact that Λ is a diagonal matrix, and using the representation of m as given by Eq. (3.8) we have derived Eq. (6.2).

Lemma 6.1 *The equation*

$$\left(\zeta^3 - \zeta^2\right)\sum_{i=1}^{p}\frac{\ell_i^2}{(\zeta+\lambda_i)^2} + \left(\zeta - \frac{1}{3}\right)\sum_{i=1}^{p}\frac{\lambda_i \ell_i^2}{(\zeta+\lambda_i)^2} = 0 \; ,$$

has a unique solution in $\left(\frac{1}{3}, 1\right)$.

Proof. It is a trivial exercise in calculus to show that

$$\frac{\zeta^2(1-\zeta)}{\zeta - \frac{1}{3}} \tag{6.15}$$

is monotone decreasing on $(\frac{1}{3}, 1]$, has a vertical asymptote at $\frac{1}{3}$, and is zero at 1. Likewise,

$$h(\zeta) := \frac{\sum \frac{\lambda_i \ell_i^2}{(\zeta+\lambda_i)^2}}{\sum \frac{\ell_i^2}{(\zeta+\lambda_i)^2}} \tag{6.16}$$

is seen to be a monotone increasing function which has a positive value at zero,

$$h(0) = \frac{\sum \frac{1}{\lambda_i}\ell_i^2}{\sum \frac{1}{\lambda_i^2}\ell_i^2} \; , \tag{6.17}$$

and which approaches the value $h(\infty) := \sum \lambda_i \ell_i^2$, when $\zeta \to \infty$. The two facts together prove Lemma 6.1.

Remark: The analysis of the function h implies the following inequality to hold in general:

$$E[Y] \leq E\left[\frac{1}{Y}\right] E\left[Y^2\right] \; , \tag{6.18}$$

where Y is any positive random variable, and E denotes the expectation operator. An independent proof of Ineq. (6.18) can be given by applying Jensen's inequality twice, first in the form $(E[x])^2 \leq E[x^2]$, and second in the form $\phi(E[x]) \leq E[\phi(x)]$, $\phi(x) = \frac{1}{x}$, $x > 0$.

The second derivation of the linear minimax estimator for ellipsoidal constraints which leads to Eq. (6.1) as well as Eq. (6.3) is based on Theorem 3.2 and the following Lemma which is also crucial for the analysis of Example 4, see below.

Lemma 6.2 *Let* $\alpha := \|\theta\|^2$, $z = z(\alpha)$ *be defined by as the unique real root of (cf.* (3.8)*)*

$$z^3 - z^2 + \frac{1}{\alpha}\left(z - \frac{1}{3}\right) = 0,$$

and let $g(\alpha) := \dfrac{z(1+\alpha z^2)}{(1+\alpha z)^3}$*. Then* $g(\alpha)$ *is a monotone decreasing function of* α*.*

Proof. We shall show that $\left(\ln(g)\right)' = \dfrac{d}{d\alpha}\ln\left(g(\alpha)\right) \leq 0$. Taking the logarithm of g we get

$$\ln(g) = \ln(z) + \ln\left(1+\alpha z^2\right) - 3\ln(1+\alpha z).$$

Implicitly differentiating $\ln(g)$ with respect to α we obtain

$$\left(\ln(g)\right)' = \frac{z'}{z} + \frac{(z^2 + 2\alpha z z')}{(1+\alpha z^2)} - \frac{3(z+\alpha z')}{(1+\alpha z)},$$

and collecting terms in z' we have

$$\left(\ln(g)\right)' = z'\left\{\frac{1}{z} + \frac{2\alpha z}{1+\alpha z^2} - \frac{3\alpha}{1+\alpha z}\right\} + \frac{z^2}{1+\alpha z^2} - \frac{3z}{1+\alpha z}.$$

Let $N := z(1+\alpha z^2)(1+\alpha z)$, then we get

$$\begin{aligned}
\left(\ln(g)\right)' &= \frac{z'}{N}\left\{(1+\alpha z^2)(1+\alpha z) + 2\alpha z^2(1+\alpha z) - 3\alpha z(1+\alpha z^2)\right\} \\
&\quad + \frac{z^2}{N}\left[z(1+\alpha z) - 3(1+\alpha z^2)\right] \\
&= \frac{z'}{N}\left\{(1+3\alpha z^2)(1+\alpha z) - 3\alpha z - 3\alpha^2 z^3\right\} \\
&\quad + \frac{z^2}{N}\left[z - 3 - 2\alpha z^2\right]. \qquad (6.19)
\end{aligned}$$

Since $\alpha(z^3 - z^2) + \left(z - \frac{1}{3}\right) \equiv 0$, we obtain by implicitely differentiating with respect to α, and collecting terms in z' the following formula for z':

$$\begin{aligned}
(z^3 - z^2) + \alpha(3z^2 z' - 2zz') + z' &= 0 \\
\Leftrightarrow \quad z'\left[3\alpha z^2 - 2z\alpha + 1\right] &= -(z^3 - z^2) \qquad (6.20) \\
\Leftrightarrow \quad z' &= -\tfrac{(z^3 - z^2)}{(3\alpha z^2 - 2\alpha z + 1)}
\end{aligned}$$

Substituting the right-hand side of Eq. (6.20) into Eq. (6.19) we obtain, after simplifying terms,

$$\Big(\ln(g)\Big)' = -\frac{z^2(z-1)}{(3\alpha z^2 - 2\alpha z + 1)} \cdot \frac{1}{N}\Big\{1 + 3\alpha z^2 + \alpha z + 3\alpha^2 z^2 - 3\alpha z - 3\alpha^2 z^2\Big\}$$

$$+\frac{z^2}{N}\left[z - 3 - 2\alpha z^2\right] \tag{6.21}$$

$$= \frac{1}{N}\left(-z^3 + z^2 + z^3 - 3z^2 - 2\alpha z^4\right) = \frac{1}{N}\left(-2z^2\right)\left(1 + \alpha z^2\right) \leq 0,$$

which concludes the proof of Lemma 6.2. $\diamond$

Using short-hand notation, the constrained optimization problem described in Theorem 3.2 is of the form

$$\sup_{\theta'\Lambda\theta\leq 1} \Big\{ \langle \ell, \theta\rangle^4 g(\|\theta\|^2) \Big\}, \tag{6.22}$$

where $g = g(\|\theta\|^2)$ is defined in Lemma 6.2.

By the necessary optimality conditions there is some number μ (below, g' denotes the derivative of g evaluated at the appropriate argument) such that

$$4g\langle \ell, \theta\rangle^3 \ell = -2g'\langle \ell, \theta\rangle^4 \left(\mathrm{id} + \frac{\mu}{2g'\langle \ell, \theta\rangle^4}\,\Lambda\right)\theta\,, \tag{6.23}$$

and

$$\mu(1 - \theta'\Lambda\theta) = 0\,. \tag{6.24}$$

Eq. (6.23) can be written as

$$-\frac{2g}{g'}\frac{\ell}{\langle \ell, \theta\rangle} = \left(id + \frac{1}{\zeta}\Lambda\right)\theta, \tag{6.25}$$

where

$$\frac{1}{\zeta} = \frac{\mu}{2g'\langle \ell, \theta\rangle^4}.$$

It follows from Lemma 3.3 that $1 = \theta'\Lambda\theta$, so we get

$$\theta = \frac{\left(\mathrm{id} + \frac{1}{\zeta}\Lambda\right)^{-1}\ell}{\sqrt{\ell'\Big(\left(\mathrm{id} + \frac{1}{\zeta}\Lambda\right)^{-1}\Big)'\Lambda\left(\mathrm{id} + \frac{1}{\zeta}\Lambda\right)^{-1}\ell}}\,. \tag{6.26}$$

Furthermore, using the formula for the derivative of $\ln(g)$ derived in the proof of Lemma 6.2, see Eq. (6.21), taking the scalar product of θ with both sides of Eq. (6.25), and using $\theta'\Lambda\theta = 1$ we obtain, $\alpha := \|\theta\|^2$,

$$\frac{1 + z\alpha}{z} = \alpha + \frac{1}{\zeta}.$$

This trivially implies that $z \equiv \zeta$. Moreover, Eq. (6.25) is equivalent to

$$\begin{aligned} \left(id + \frac{1}{\varrho}\Lambda\right)^{-1} \ell &= -\frac{g'}{2g}\langle \ell, \theta\rangle\theta \\ &= \frac{z}{1+z\|\theta\|^2}\langle \ell, \theta\rangle\theta = m, \end{aligned}$$

by Eq. (3.20).

So we have seen that both optimization problems yield the same answer, as of course they should do. The difference between the two approaches is that for the inf-sup problem the Eq. (6.14) needs to be analyzed which is actually a more complicated equation than Eq. (3.8). On the other hand, before one can solve Eq. (3.8) one needs to know $\|\theta^*\|^2$, which requires solving the non-linear optimization problem (4.2).

Example 3. Let Y be given as described by Eq. (1.1), and $\theta \in \Theta \subset L_2[0,1]$, where

$$\Theta = \left\{\theta \middle| D_\alpha \theta = 0,\ \sup_{0\le t\le 1} |\theta(t)| \le B\right\}$$

for some fixed constant B, and D_α is the differential operator

$$D_\alpha \theta(t) = \frac{d}{dt}\theta(t) + \alpha\theta(t).$$

Consider the problem of estimating $L(\theta) = \theta(0)$, i.e. estimate the initial value of the unknown parameter function. Now, every member of Θ has the representation $\theta(t) = Ce^{-\alpha t}$ for some C with $|C| \le B$. Therefore, the set Θ is closed in the L_2 topology and it is clearly bounded. Hence, Θ is weakly compact. Moreover, since $\int_0^1 \theta(s)ds = C\left(\frac{1-e^{-\alpha}}{\alpha}\right)$, we have $L(\theta) = \frac{\alpha}{1-e^{-\alpha}}\int_0^1 \theta(s)ds$, and consequently, L is weakly continuous. Therefore, Theorem 4.1 applies and

$$\sup\inf R(\theta, m) = 3\sup_{\theta\in\Theta}\left\{L\left(\frac{\theta}{\|\theta\|}\right)^4 \left(\frac{z_\theta\|\theta\|^2}{1+z_\theta\|\theta\|^2}\right)^3 \left(\frac{1+z_\theta^2\|\theta\|^2}{z_\theta^2}\right)\right\}.$$

Now,

$$\|\theta\|^2 = \int_0^1 \theta^2(s)ds = C^2\int_0^1 e^{-2ds}ds = C^2\frac{1-e^{-2\alpha}}{2\alpha}.$$

Hence,

$$\sup_{\theta\in\Theta}\left\{L\left(\frac{\theta}{\|\theta\|}\right)^4 \left(\frac{z_\theta\|\theta\|^2}{1+z_\theta\|\theta\|^2}\right)^3 \left(\frac{1+z_\theta^2\|\theta\|^2}{z_\theta^2}\right)\right\} = \sup_{|c|\le B}\left\{\left(\frac{1-e^{-2\alpha}}{2\alpha}\right)^{-2}\cdot\Psi(C)\right\},$$

where $\Psi(C) := \left(\frac{z_\theta\|\theta\|^2}{1+z_\theta\|\theta\|^2}\right)^3 \left(\frac{1+z_\theta^2\|\theta\|^2}{z_\theta^2}\right)$, $\theta(t) = Ce^{-\alpha t}$, and, by Lemma 6.2, Ψ is monoton increasing in C.

Thus, the supremum is attained at $C = \pm B$, and the corresponding solution is $\theta^*(t) = \pm B e^{-\alpha t}$. Therefore the 'minimax' weight function m^* is given by (cf. (4.3))

$$m^*(t) = \frac{z^* B^2}{1 + z^* B^2 \frac{(1-e^{-2\alpha})}{2\alpha}}\, e^{-\alpha t}, \tag{6.27}$$

where z^* is the solution of:

$$z^3 - z^2 + \frac{2\alpha}{B^2\,(1-e^{-2\alpha})}\left(z - \frac{1}{3}\right) = 0, \tag{6.28}$$

and the linear minimax estimate of $\theta(0)$ equals

$$\widehat{\theta(0)} = \frac{z^* B^2}{1 + z^* B^2 \left(\frac{1-e^{-2\alpha}}{2\alpha}\right)} \int_0^1 e^{-\alpha t} dY(t). \tag{6.29}$$

Example 4. The problem is to linearly estimate the initial value of $\theta(t)$ given the data $\left(Y(t)\right)_{0\le t\le 1}$ when it is known that $\theta(t) = x + \varrho(t)t$, where $|\varrho(t)| \le B$, B a known constant, and $x \in \mathbb{R}$. This problem has been analyzed for the quadratic loss function in [6], cf. also [9]. Here we consider the case of the quartic loss function. To exclude trivial cases we assume $|x| \le B$ and $B^2 \ge 18$.

The linear mini-max estimator for the case under consideration is determined by the solution of the optimization problem:

$$\max_{\theta\in\Theta} \left\{ L(\theta)^4 g\left(\|\theta\|^2\right) \right\}.$$

Since $\theta \in \Theta \Leftrightarrow \theta(t) = x + t\varrho(t)$, w.l.o.g. $x \in \mathbb{R}^+$. Since $L(\theta) = x$, this optimization problem is equivalent to

$$\max_{x>0} \left\{ x^4 \max_{\varrho, |\varrho|\le B} \left\{ g\left(\|\theta\|^2\right) \right\} \right\}. \tag{6.30}$$

Since g is monotone decreasing the maximization problem involving ϱ is equivalent to:

$$\min_{\varrho,|\varrho|\le B} \left\{ \|\theta\|^2 \right\} = \min_{\varrho,|\varrho|\le B} \left\{ \|x + t\varrho(t)\|^2 \right\}.$$

The solution to this minimization problem is given by:

$$\varrho^*(t) = \begin{cases} -B \quad , & 0 \le t \le \frac{x}{B} \\ -\frac{x}{t} \quad , & \frac{x}{B} \le t \le 1 \end{cases} \tag{6.31}$$

The square of the L_2-norm of $\theta^*(t) = x + t\varrho^*(t)$ is therefore easily computed to be:

$$\|\theta^*\|^2 = \frac{1}{3}\frac{x^3}{B}. \tag{6.32}$$

Substituting the right-hand side of Eq. (6.32) into the x-optimization problem described by (6.30) and using $\alpha = \frac{1}{3}\frac{x^3}{B}$, the x-optimization problem reads:

$$\max_{\alpha>0}\Big\{f(\alpha)\Big\},$$

where $f(\alpha) := \text{const}\,\alpha^{\frac{4}{3}} g(\alpha),\ \text{const} = (3B)^{\frac{4}{3}}$.

Exploiting the formula for the derivative of g, proved in Lemma 6.2, we find the stationary points of f to be characterized by the equation

$$2 = \alpha z(\alpha) \Leftrightarrow 2 = \frac{1}{3}\frac{x^3}{B} z. \tag{6.33}$$

This equation together with the identity for z (as a function of $\alpha = \frac{1}{3}\frac{x^3}{B}$) yields

$$z^* = z(\alpha) = \frac{1}{4} + \sqrt{\frac{11}{48}}. \tag{6.34}$$

Thus

$$x^* = \left(\frac{24B}{1+\sqrt{\frac{11}{3}}}\right)^{\frac{1}{3}},$$

and the "best" linear estimate of the initial value equals:

$$\begin{aligned}
\hat{x} &= (x^*)^2 \frac{z^*}{1 + z^*\left(\frac{8}{1+\sqrt{\frac{11}{3}}}\right)} Y(1) - x^* \frac{z^*}{1 + z^*\left(\frac{8}{1+\sqrt{\frac{11}{3}}}\right)} \int_0^1 t\varrho^*(t)dY_t \\
&= \frac{x^* z^*}{3}\left[x^* Y(1) - \int_0^1 t\varrho^*(t)dY_t\right],
\end{aligned}$$

where ϱ^* is defined by (6.31) with x replaced by x^*.

Finally, the value for z^* as given by Eq. (6.34) is consistent with the vaule of the solution of

$$z^3 - z^2 + \left(\frac{1+\sqrt{\frac{11}{3}}}{8}\right)\left(z - \frac{1}{3}\right) = 0,$$

viz.

$$z^* \approx 0.7287\,.$$

7 Summary

We have dealt with the following statistical estimation problem:

$$\min_{m \in L_2[0,1]} \max_{\theta \in \Theta} R(\theta, m), \quad \text{where} \quad R(\theta, m) = E\left[\left(L(\theta) - \langle m, Y\rangle\right)^4\right],$$

where Y is given by Eq. (1.1). Using convex analysis we have shown that under appropriate conditions on Θ a mini-max Theorem holds (see Theorem 3.1 for the finite-dimensional case and Theorem 4.1 for the infinite-dimensional case). The mini-max result leads to a characterization of the linear mini-max estimator in terms of the solution of a non-linear optimization problem, see Theorems 3.2 and 4.1. We use the characterization to explicitely solve specific finite-dimensional as well as infinite-dimensional examples, e.g. the case when Θ is an ellipsoid, see Example 2. Example 2 (see also Example 1 and Eq. (3.9)) clearly shows the effect the loss function has on the form of the estimator. In the case of *quartic* loss function, when compared with the *quadratic* loss function, the shrinkage factor is modified and depends in a highly non-linear way on $\|\theta\|^2$. Phrased in a "loose way", the estimate is "pulled closer to the origin". Of course, this fact fits one's intuition.

References

[1] Alpern, S. and Gal, S. (1988). A mixed-strategy minimax theorem without compactness. *SIAM J. Control and Optimization,* vol. 26, No. 6, 1357–1361.

[2] Drygas, H. (1991). On an extension of the Girko inequality in linear minimax estimation. Proceeding of the Probastat '91 Conference, Bratislava, CSFR.

[3] Helmes, K. and Christopeit, N. (1996). Linear estimation with ellipsoidal constraints. *Acta Applicandae Mathematicae,* vol. 43, No. 1, 3–15.

[4] Helmes, K. and Srinivasan, C. (1992). Linear minimax estimation. *Lecture Notes in Economics and Mathematics,* 389, 9–23.

[5] Helmes, K. and Srinivasan, C. (1996). Computing linear mini-max estimators. Proceeding of the 2nd International Conference on Game Theory and its Applications, Bangalore, India.

[6] Ibragimov, A.D. and Hasminski, R.Z. (1984). On non-parametric estimation of the value of a linear function in Gaussian white noise. *Theory of Probability and Its Applications,* Vol. XXIX, No. 1, 18–32.

[7] Ibragimov, A.D. and Hasminski, R.Z. (1987). Estimation of linear functionals in Gaussian noise. *Theory Prob. Appl.,* vol. 32, No. 1, 30–39.

[8] Ioffe, A.D. and Levin, V.L. (1974). Subdifferentials of convex functions. *Transactions of the Moscow Mathematical Society,* vol. 26.

[9] Legostaeva, I.L. and Shiryaev, A.N. (1971). Minimax weights in a trend detection problem of a random process, *Theory Prob. Appl.,* vol. 16, 344–349.

[10] Moreau, J.J. (1964). Sur la fonction polair dúne fonction semi-continue supervieurment, *C.R. Acad. Sci. Paris,* 258, 1128–1130.

[11] Nussbaum, M. (1995). Asymptotic Equivalence of Density Estimation and Gaussian White Noise. Preprint WIAS No. 174.

[12] Pinelis, I.F. (1988). On minimax risk. *Theory Prob. Appl.,* vol. 33, 104–109.

[13] Stahlecker, P. and Drygas, H. (1992). Representation theorems in linear minimax estimation. Report No. V-85-92, University of Oldenburg.