1. ESTIMATION OF LINEAR FORMS

We consider the linear regression model

$$y = X\beta + \varepsilon \tag{1.1}$$

where y is the $n \times 1$ – vector of observations, X a nonrandom $n \times K$ – regression matrix, β an unknown K–dimensional parameter vector and ε the $n \times 1$–vector of disturbances. In contrast to conventional econometric theory, we shall treat the disturbance vector as an additonal unknown parameter vector. One objective in this section is to find linear estimates of a linear form $a^{\mathrm{T}}\beta$ which are "best" in a sense defined below. We assume that we are given the a priori information

$$\|\beta\| \leq 1, \qquad \|\varepsilon\| \leq 1.$$

Then in the class of all estimates of the form $\widetilde{a^{\mathrm{T}}\beta} = b^{\mathrm{T}}y$ (b a nonrandom $n \times 1$–vector), we want to find the linear *minimax* estimate $\overline{a^{\mathrm{T}}\beta} = \bar{b}^{\mathrm{T}}y$, i.e. that linear form $\bar{b}$ for which

$$\min_b \max_{\|\beta\|,\|\varepsilon\| \leq 1} [a^{\mathrm{T}}\beta - b^{\mathrm{T}}y]^2 = \max_{\|\beta\|,\|\varepsilon\| \leq 1} [a^{\mathrm{T}}\beta - \bar{b}^{\mathrm{T}}y]^2. \tag{1.2}$$

In order to solve this problem, note that

$$\begin{aligned}[a^{\mathrm{T}}\beta - b^{\mathrm{T}}y]^2 &= [(a - Xb)^{\mathrm{T}}\beta - b^{\mathrm{T}}\varepsilon]^2 \\ &= (c^{\mathrm{T}}\beta)^2 + (b^{\mathrm{T}}\varepsilon)^2 - 2(c^{\mathrm{T}}\beta)(b^{\mathrm{T}}\varepsilon),\end{aligned}$$

where we have put $c = a - X^{\mathrm{T}}b$. Evidently, the maximum over $\|\beta\|, \|\varepsilon\| \leq 1$ is attained at, e.g., $\beta = c/\|c\|, \varepsilon = -b/\|b\|$, with

$$\max_{\|\beta\|,\|\varepsilon\| \leq 1} [a^{\mathrm{T}}\beta - b^{\mathrm{T}}y]^2 = (\sqrt{c^{\mathrm{T}}c} + \sqrt{b^{\mathrm{T}}b})^2.$$

So we have to solve the minimization problem

$$f(b) = \sqrt{(a - X^{\mathrm{T}}b)^{\mathrm{T}}(a - X^{\mathrm{T}}b)} + \sqrt{b^{\mathrm{T}}b} \to \min_b! \tag{1.3}$$

Case 1. Either $\lambda_{\min}(X^{\mathrm{T}}X) > 1$ or $\lambda_{\max}(X^{\mathrm{T}}X) < 1$.
Suppose that the minimum is attained at some point $\bar{b}$ satisfying $\bar{b} \neq 0$ and $a - X^{\mathrm{T}}\bar{b} \neq 0$. Then f is differentiable at $\bar{b}$, and the first order necessary optimality condition is

$$\nabla f(\bar{b}) = \frac{2X(X^{\mathrm{T}}\bar{b} - a)}{\sqrt{(a - X^{\mathrm{T}}\bar{b})^{\mathrm{T}}(a - X^{\mathrm{T}}\bar{b})}} + \frac{2\bar{b}}{\sqrt{\bar{b}^{\mathrm{T}}\bar{b}}} = 0$$

or

$$\frac{\bar{c}^{\mathrm{T}}X^{\mathrm{T}}}{\sqrt{\bar{c}^{\mathrm{T}}\bar{c}}} = \frac{\bar{b}^{\mathrm{T}}}{\sqrt{\bar{b}^{\mathrm{T}}\bar{b}}}, \tag{1.4}$$

Translated by A. I. Vladimirova and N. Christopeit with $\bar{c} = a - X^{\mathrm{T}}\bar{b}$. (1.4) implies that

$$\frac{\bar{c}^{\mathrm{T}} X^{\mathrm{T}} X \bar{c}}{\bar{c}^{\mathrm{T}} \bar{c}} = 1.$$

As a consequence,

$$\lambda_{\min}(X^{\mathrm{T}}X) \leq 1 \leq \lambda_{\max}(X^{\mathrm{T}}X),$$

contradicting our assumption. Hence the only candidates for a minimum are $\bar{b} = 0$ and any $\bar{b}$ satisfying $\bar{c} = a - X^{\mathrm{T}}\bar{b} = 0$.

a) If $a \notin \mathcal{R}(X^{\mathrm{T}})$, then $\bar{b} = 0$ is the only solution, with value

$$f(0) = \sqrt{a^{\mathrm{T}}a}.$$

Note that, in this case, $rg(X^{\mathrm{T}}X) < K$ and therefore $\lambda_{\min}(X^{\mathrm{T}}X) = 0$.

b) Otherwise, $\bar{b} = X^{+\mathrm{T}}a$ is the minimum norm solution of the equation $X^{\mathrm{T}}\bar{b} = a$, and

$$f(\bar{b}) = \sqrt{a^{\mathrm{T}}X^{+}X^{+\mathrm{T}}a}. \tag{1.5}$$

Since $X^{+}X^{+\mathrm{T}} = (X^{\mathrm{T}}X)^{+}$, the nonnull eigenvalues of $X^{+}X^{+\mathrm{T}}$ are just the inverses of the nonnull eigenvalues of $X^{\mathrm{T}}X$. Therefore,

$$\lambda_{\min}(X^{\mathrm{T}}X) > 1 \Rightarrow \lambda_{\max}(X^{+}X^{+\mathrm{T}}) < 1 \Rightarrow \sqrt{a^{\mathrm{T}}a} > \sqrt{a^{\mathrm{T}}X^{+}X^{+\mathrm{T}}a}. \tag{1.6}$$

To deal with the case $\lambda_{\max}(X^{\mathrm{T}}X) < 1$, let $rg(X^{\mathrm{T}}X) = p \leq K$ and consider the diagonalization

$$X^{\mathrm{T}}X = T\Lambda T^{\mathrm{T}},$$

where $\Lambda = diag(\lambda_1, \ldots, \lambda_p, 0, \ldots, 0)$, λ_i are the nonnull eigenvalues of $X^{\mathrm{T}}X$, and T is an orthogonal matrix. Then

$$X^{+}X^{+\mathrm{T}} = T\Lambda^{+}T^{\mathrm{T}}$$

with $\Lambda^{+} = \mathrm{diag}(\lambda_1^{-1}, \ldots, \lambda_p^{-1}, 0, \ldots, 0)$. Since $\mathcal{R}(X^{\mathrm{T}}) = \mathcal{R}(X^{\mathrm{T}}X)$, $a \in \mathcal{R}(X^{\mathrm{T}})$ implies that

$$a = X^{\mathrm{T}}Xw = T\Lambda T^{\mathrm{T}}w$$

for some w and hence

$$z = T^{\mathrm{T}}a = \Lambda T^{\mathrm{T}}w = \Lambda v.$$

This means that the last $K - p$ components of z must be zero. Consequently, since

$$\lambda_{\max}(X^{\mathrm{T}}X) < 1 \Rightarrow \lambda_i^{-1} > 1 \text{ for all } i = 1, \ldots, p,$$

we find that

$$a^{\mathrm{T}}a = z^{\mathrm{T}}z = \sum_{i=1}^{p} z_i^2 < \sum_{i=1}^{p} \lambda_i^{-1} z_i^2 = z^{\mathrm{T}}\Lambda^{+}z = a^{\mathrm{T}}X^{+}X^{+\mathrm{T}}a,$$

hence

$$\lambda_{\max}(X^{\mathrm{T}}X) < 1 \Rightarrow \sqrt{a^{\mathrm{T}}a} < \sqrt{a^{\mathrm{T}}X^{+}X^{+\mathrm{T}}a}. \tag{1.7}$$

Gathering the results of a) and b) (1.6), (1.7), we arrive at the following result.

PROPOSITION 1. *In case 1, the minimax solution is given by*

$$\bar{b} = \begin{cases} X^{+\mathrm{T}}a & if \quad \lambda_{\min}(X^{\mathrm{T}}X) > 1 \\ 0 & if \quad \lambda_{\max}(X^{\mathrm{T}}X) < 1. \end{cases}$$

Note that, in the first case, the *minimax* estimate of $a^{\mathrm{T}}\beta$ is given by

$$a^{\hat{\mathrm{T}}}\beta = \bar{b}^{\mathrm{T}}y = a^{\mathrm{T}}X^{+}y = a^{\mathrm{T}}\hat{\beta},$$

where $\hat{\beta}$ is the ordinary least squares estimate of β.

Case 2. $\lambda_{\min}(X^{\mathrm{T}}X) \leq 1 \leq \lambda_{\max}(X^{\mathrm{T}}X)$.

If the minimum is attained at some b s.t. $b \neq 0, c \neq 0$ equation (1.4) must be satisfied. Denoting $r = \sqrt{c^{\mathrm{T}}c}/\sqrt{b^{\mathrm{T}}b}$, $c = a - X^{\mathrm{T}}b$,

$$X(a - X^{\mathrm{T}}b) = rb$$

or

$$(XX^{\mathrm{T}} + rI)b = Xa.$$

Apparently, $M(r) = XX^{\mathrm{T}} + rI$ is positive definite for all $r > 0$, hence

$$b = M(r)^{-1}Xa. \tag{1.8}$$

Simultaneously, since $c = c(r) = (I - X^{\mathrm{T}}M(r)^{-1}X)a$,

$$a^{\mathrm{T}}[I - X^{\mathrm{T}}M(r)^{-1}X]^2 a = r^2 a^{\mathrm{T}}X^{\mathrm{T}}M(r)^{-2}Xa \tag{1.9}$$

should be satisfied by definition of r. Let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_k \geq 0$ be the eigenvalues of $X^{\mathrm{T}}X$. Then XX^{T} can be diagonalized in the form

$$XX^{\mathrm{T}} = \bar{T}\bar{\Lambda}\bar{T}^{\mathrm{T}}, \tag{1.10}$$

where $\bar{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_K, 0, \ldots, 0)$ (with $n - K$ zeroes) is the diagonal matrix of eigenvalues of XX^{T} and $\bar{T}$ has as its columns orthonormal eigenvectors of XX^{T}. As a consequence, for integer k,

$$\begin{aligned} M^k &= \bar{T}(\bar{\Lambda} + rI)^k\bar{T}^{\mathrm{T}}, \\ XX^{\mathrm{T}}M^k &= \bar{T}\bar{\Lambda}(\bar{\Lambda} + rI)^k\bar{T}^{\mathrm{T}}, \end{aligned} \tag{1.11}$$

i.e. $XX^{\mathrm{T}}M^k$ has eigenvalues $\lambda_i(\lambda_i + r)^k$, $i = 1, \ldots, K$, and 0 ($n - K$ times). Moreover,

$$X^{\mathrm{T}}M^kXw = \lambda w \Rightarrow XX^{\mathrm{T}}M^kXw = \lambda Xw,$$

i.e. every eigenvalue of $X^{\mathrm{T}}M^kX$ is an eigenvalue of $XX^{\mathrm{T}}M^k$. In particular, the nonnull eigenvalues of $X^{\mathrm{T}}M^kX$ are given by $\lambda_i(\lambda_i + r)^k$ for $\lambda_i > 0$. Finally, it follows easily from

(1.10) and (1.11) that $X^{\mathrm{T}}M^kX$ and $X^{\mathrm{T}}M^lX$ commute for all pairs (k,l) of integers and can therefore be simultaneously diagonalized. Hence, in particular,

$$X^{\mathrm{T}}M^{-1}X = T\Lambda(\Lambda + rI)^{-1}T^{\mathrm{T}}, \quad X^{\mathrm{T}}M^{-2}X = T\Lambda(\Lambda + rI)^{-2}T^{\mathrm{T}}$$

with $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_K)$ and for some orthogonal matrix T, and

$$\begin{aligned}[I - X^{\mathrm{T}}M^{-1}X]^2 &= T[I - \Lambda(\Lambda + rI)^{-1}]^2 T^{\mathrm{T}} \\ &= T\mathrm{diag}\left(\frac{r^2}{(\lambda_1 + r)^2}, ..., \frac{r^2}{(\lambda_k + r)^2}\right) T^{\mathrm{T}}.\end{aligned}$$

Therefore, denoting $z = T^{\mathrm{T}}a$, (1.9) becomes

$$\sum_{i=1}^{K} \frac{r^2 z_i^2}{(\lambda_i + r)^2} = r^2 \sum_{i=1}^{K} \frac{\lambda_i z_i^2}{(\lambda_i + r)^2}$$

or

$$\sum_{i=1}^{K} \frac{\lambda_i - 1}{(\lambda_i + r)^2} z_i^2 = 0. \tag{1.12}$$

In a typical nontrivial case, where there are $l > 0$ eigenvalues < 1 $(1 \leq l < K)$, m eigenvalues $= 1$ $(0 \leq m < K - l)$ and $K - l - m > 0$ eigenvalues > 1:

$$0 \leq \lambda_K \leq \lambda_{K-1} \leq ... \leq \lambda_{K-l+1} < 1 = \lambda_{K-l} = ... = \lambda_{K-l-m+1} < \lambda_{K-l-m} \leq ... \leq \lambda_1,$$

(1.12) becomes

$$\begin{aligned}&\prod_{j=K-l+1}^{K} (\lambda_j + r)^2 \sum_{i=1}^{K-l-m} c_i \prod_{j=1, j\neq i}^{K-l-m} (\lambda_j + r)^2 \\ &\qquad - \prod_{j=1}^{K-l-m} (\lambda_j + r)^2 \sum_{i=K-l+1}^{K} c_i \prod_{j=K-l+1, j\neq i}^{K} (\lambda_j + r)^2 = p(r) - q(r) = 0,\end{aligned} \tag{1.13}$$

with nonnegative coefficients c_i. Both $p(r)$ and $q(r)$ are polynomials of degree $\leq 2(K - m - 1)$ and nonnegative for all r. Depending on the constellation of parameters, there may be from 0 to $2(k - m) - 1$ positive roots of equation (1.13) (there always exists at least one negative root).

For every positive root r we have to calculate $f(b)$ with $b = b(r)$ given by (1.8):

$$\begin{aligned}g(r) = f(b(r)) &= \sqrt{c(r)^{\mathrm{T}}c(r)} + \sqrt{b(r)^{\mathrm{T}}b(r)} \\ &= (1 + r)\sqrt{b(r)^{\mathrm{T}}b(r)} \\ &= (1 + r)\sqrt{a^{\mathrm{T}}X^{\mathrm{T}}M(r)^{-2}Xa} \\ &= (1 + r)\sqrt{\sum_{i=1}^{K} \frac{\lambda_i z_i^2}{(\lambda_i + r)^2}} \\ &= (1 + r)\sqrt{\sum_{i=1}^{p} \frac{\lambda_i z_i^2}{(\lambda_i + r)^2}},\end{aligned} \tag{1.14}$$

if $\lambda_K = ... = \lambda_{p+1} = 0$, and compare these values with

$$f(0) = \sqrt{a^{\mathrm{T}}a} = \sqrt{z^{\mathrm{T}}z} \quad \text{and} \quad f(\bar{b}) = \sqrt{a^{\mathrm{T}}X^{+}X^{+^{\mathrm{T}}}a}.$$

In doing so, note that

$$\lim_{r\downarrow 0} M(r)^{-1}X = X^{+\mathrm{T}},$$

hence $\bar{b} = b(0+)$ (for $b(r)$ given by (1.8)) and

$$f(\bar{b}) = g(0) = \sqrt{\sum_{i=1}^{p} \lambda_i^{-1} z_i^2}.$$

Moreover, for any positive root r of (1.12),

$$\begin{aligned}
f(0)^2 - g(r)^2 &= \sum_{i=1}^{K} z_i^2 - \sum_{i=1}^{K} \frac{(1+r)^2\lambda_i z_i^2}{(\lambda_i + r)^2} \\
&= \sum_{i=1}^{K} \frac{\lambda_i(\lambda_i - 1) + r^2(1-\lambda_i)}{(\lambda_i + r)^2} z_i^2 \\
&= \sum_{i=1}^{K} \lambda_i \frac{\lambda_i - 1}{(\lambda_i + r)^2} z_i^2 > \sum_{i=1}^{K} \frac{\lambda_i - 1}{(\lambda_i + r)^2} z_i^2 = 0,
\end{aligned}$$

where the third equality follows from (1.12) and the strict inequality is valid for the nontrivial case where not all $z_i, i = 1, \ldots, K-l-m$, and not all $z_i, i = K-l+1, \ldots, K$, are equal to zero. Hence, in this case, if there exists at least one positive root r, $b = 0$ can be excluded from the candidates for optimal linear combinations. Similarly, if all $\lambda_i > 0$, for any positive root r of (1.12)

$$\begin{aligned}
g(0)^2 - g(r)^2 &= \sum_{i=1}^{K} \frac{z_i^2}{\lambda_i} - (1+r)^2 \sum_{i=1}^{K} \frac{\lambda_i z_i^2}{(\lambda_i + r)^2} \\
&= \sum_{i=1}^{K} \frac{z_i^2}{\lambda_i} - (1+r)^2 \sum_{i=1}^{K} \frac{z_i^2}{(\lambda_i + r)^2} \\
&= \sum_{i=1}^{K} \frac{r^2(1-\lambda_i) + \lambda_i(\lambda_i - 1)}{\lambda_i(\lambda_i + r)^2} z_i^2 \\
&= r^2 \sum_{i=1}^{K} \frac{1-\lambda_i}{\lambda_i(\lambda_i + r)^2} z_i^2 + \sum_{i=1}^{K} \frac{\lambda_i - 1}{(\lambda_i + r)^2} z_i^2 \\
&= r^2 \sum_{i=1}^{K} \frac{1-\lambda_i}{\lambda_i(\lambda_i + r)^2} z_i^2 > r^2 \sum_{i=1}^{K} \frac{1-\lambda_i}{(\lambda_i + r)^2} z_i^2 = 0,
\end{aligned}$$

where again (1.12) has been used for the second, fifth and sixth equality. Hence, in the nontrivial case as described above, if all λ_i are positive, $\bar{b} = b(0+)$, too, does not qualify as a candidate for the minimum of (1.3) in case there exists at least one positive root of (1.12).

Denoting $\mathcal{R}^+$ the set of positive roots of (1.12), we thus arrive at

PROPOSITION 2. *In case 2, in a nontrivial situation (as described above) and if $\mathcal{R}^+ \neq \emptyset$, the absolute minimum of (1.3) is provided by*

$$\operatorname{argmin}\{f(b(r)) : r \in \mathcal{R}^+\} \tag{1.15}$$

if all λ_i are positive. If some λ_i are zero, choice must be made between (1.14) and $\bar{b}$. In the case where $\mathcal{R}^+ = \emptyset$, choice must be made between $b = 0$ and $b = \bar{b}$.

Hence in case 2, there seems to be no nice closed form solution as in case 1.

2. ESTIMATION OF PARAMETERS UNDER CIRCULAR CONSTRAINTS

The setting in this section will again be the linear model (1.1) considered in Section 1, together with the constraints

$$\|\beta\| \leq 1, \qquad \|\varepsilon\| \leq 1.$$

Our objective now is, however, to find the *linear minimax estimator* of the parameter vector β itself, i.e. among all estimates of the form $\tilde{\beta} = Cy$ we want to find the one $\bar{\beta} = \bar{C}y$ for which

$$\min_{C \in \mathbb{R}^{K \times n}} \max_{\|\beta\|, \|\varepsilon\| \leq 1} \|\beta - Cy\|^2 = \max_{\|\beta\|, \|\varepsilon\| \leq 1} \|\beta - \bar{C}y\|^2. \tag{2.1}$$

Inserting y from (1.1) and denoting

$$\begin{aligned} F(C; \beta, \varepsilon) &= \|\beta - Cy\|^2 \\ &= \beta^{\mathrm{T}}(I - CX)^{\mathrm{T}}(I - CX)\beta - 2\beta^{\mathrm{T}}(I - CX)^{\mathrm{T}}C\varepsilon + \varepsilon^{\mathrm{T}}C^{\mathrm{T}}C\varepsilon \\ \bar{F}(C) &= \max_{\|\beta\|, \|\varepsilon\| \leq 1} F(C; \beta, \varepsilon), \end{aligned}$$

(2.1) becomes the minimization problem

$$\bar{F}(C) \longrightarrow \min_{C \in \mathbb{R}^{K \times n}}!$$

Note that, for fixed $\beta, \varepsilon, F(C; \beta, \varepsilon)$ is convex in C and hence so is $\bar{F}(C)$. For any convex functional $f(C)$, let $\partial f(C_0)$ denote the *subdifferential* at C_0, i.e. the set of all $\Phi \in \mathbb{R}^{K \times n}$ s.t.

$$\langle \Phi, C - C_0 \rangle \leq f(C) - f(C_0)$$

for all $C \in \mathbb{R}^{K \times n}$, where $\langle \cdot, \cdot \rangle$ denotes the scalar product

$$\langle \Phi, C \rangle = \operatorname{Tr} \Phi^{\mathrm{T}} C.$$

Since $\bar{F}$ is finite and hence continuous, $\partial \bar{F}(C)$ is nonvoid for all C, and we have the following elementary optimality condition.

LEMMA 1. *$\bar{C}$ is a minimal point of $\bar{F}$ $\iff$ $0 \in \partial \bar{F}(\bar{C})$.*

The basis for the evaluation of $\bar{F}(\bar{C})$ is the following result.

LEMMA 2. (Joffe and Levin). *Let T, Z be topological vector spaces, $F : Z \times T \to I\!R$ a continuous functional s.t., for each t, $f(t, z)$ is convex in z. Assume that T is compact and let*

$$\bar{f}(z) = \max_{t \in T} f(z, t).$$

Then

$$\partial \bar{f}(z) = \bigcup_{\pi \in \Pi(z)} \int \partial_z f(z, t) \pi(\mathrm{d}t), \tag{2.2}$$

where

$$\Pi(z) = \{\pi : \pi \text{ probability measure on } T \text{ s.t. } \quad supp(\pi) \subset T^*(z)\},$$
$$T^*(z) = \{t \in T : f(z, t) = \bar{f}(z)\}.$$

For a proof and the precise meaning of the integal in (2.2), cf. Joffe and Levin [3]. Actually, in our case, the $\partial_z f(z, t)$ will be singletons, so that the meaning is clear. The situation to which we shall apply Lemma 2 is the following.

— $T = \left\{ \begin{pmatrix} \beta \\ \varepsilon \end{pmatrix} : \beta \in I\!R^K, \varepsilon \in I\!R^n, \|\beta\|, \|\varepsilon\| \leq 1 \right\}, \quad Z = I\!R^{K \times n};$
— $f(z, t) = F(C; \beta, \varepsilon);$
— $\bar{f}(z) = \bar{F}(C);$
— $\partial_C F(C; \beta, \varepsilon) = \{2[(CX - I)\beta\beta^{\mathrm{T}} X^{\mathrm{T}} + C(\varepsilon\beta^{\mathrm{T}} X^{\mathrm{T}} + X\beta\varepsilon^{\mathrm{T}}) - \beta\varepsilon^{\mathrm{T}} + C\varepsilon\varepsilon^{\mathrm{T}}]\};$
— $T^*(C) = \left\{ \begin{pmatrix} \beta \\ \varepsilon \end{pmatrix} \in T : F(C; \beta; \varepsilon) = \bar{F}(C) \right\}.$

Hence we have the equivalence

$$0 \in \partial \bar{F}(\bar{C}) \iff \text{there exists a probability measure } \pi \text{ on } T^*(\bar{C}) \text{ s.t.}$$

$$\int_{T^*(\bar{C})} [(\bar{C}X - I)\beta\beta^{\mathrm{T}} X^{\mathrm{T}} + \bar{C}(\varepsilon\beta^{\mathrm{T}} X^{\mathrm{T}} + X\beta\varepsilon^{\mathrm{T}}) - \beta\varepsilon^{\mathrm{T}} + \bar{C}\varepsilon\varepsilon^{\mathrm{T}}]\pi(d\beta, d\varepsilon) = 0. \tag{2.3}$$

We are now in the position to prove

PROPOSITION 3. *Suppose that $\lambda_{\min}(X^{\mathrm{T}} X) \geq 1$. Then the linear minimax estimator of β is given by*

$$\bar{\beta} = X^+ y.$$

Proof. Our conjecture is that, for $\bar{C} = X^+$, $0 \in \partial \bar{F}(\bar{C})$. Since $X^{\mathrm{T}} X$ is nonsingular, $\bar{C}X - I = 0$ and hence

$$F(\bar{C}; \beta, \varepsilon) = \varepsilon^{\mathrm{T}} X^{+\mathrm{T}} X^+ \varepsilon,$$
$$\bar{F}(\bar{C}) = \lambda_{\max}(X^{+\mathrm{T}} X^+) = \lambda_{\min}^{-1}(X^{\mathrm{T}} X) =: \mu.$$

Here we have used the fact that $X^{+\mathrm{T}}X^{+} = (XX^{\mathrm{T}})^{+}$ and that the nonnull eigenvalues of $(XX^{\mathrm{T}})^{+}$ are just the inverses of the nonnull eigenvalues of XX^{T}, which in term coincide with those of $X^{\mathrm{T}}X$. As a consequence,

$$T^*(\bar{C}) = \left\{ \binom{\beta}{\varepsilon} : \|\beta\| \leq 1; \varepsilon \text{ unit length eigenvector of } X^{+\mathrm{T}}X^{+} \right.$$
$$\left. \text{corresponding to eigenvalue } \mu \right\}$$

and (2.3) becomes

$$\begin{aligned} &\int_{T^*(\bar{C})} [X^{+}\varepsilon\beta^{\mathrm{T}}X^{\mathrm{T}} + X^{+}X\beta\varepsilon^{\mathrm{T}} + X^{+}\varepsilon\varepsilon^{\mathrm{T}} - \beta\varepsilon^{\mathrm{T}}]\pi(\mathrm{d}\beta, \mathrm{d}\varepsilon) \\ &= \int_{T^*(\bar{C})} X^{+}(\varepsilon\beta^{\mathrm{T}}X^{\mathrm{T}} + \varepsilon\varepsilon^{\mathrm{T}})\pi(\mathrm{d}\beta, \mathrm{d}\varepsilon) = 0. \end{aligned} \tag{2.4}$$

Let now $\bar{\varepsilon}$ be any unit length eigenvector of $X^{+\mathrm{T}}X^{+}$ coresponding to eigenvalue μ, and put

$$\bar{\beta} = -X^{+}\bar{\varepsilon}.$$

Then

$$\|\bar{\beta}\|^2 = \bar{\varepsilon}^{\mathrm{T}}X^{+\mathrm{T}}X^{+}\bar{\varepsilon} = \mu \leq 1,$$

i.e. $\binom{\bar{\beta}}{\bar{\varepsilon}} \in T^*(\bar{C})$. Let $\bar{\pi}$ denote the unit mass at $\binom{\bar{\beta}}{\bar{\varepsilon}}$. Then the second integral in (2.4) becomes

$$X^{+}(\bar{\varepsilon}\bar{\beta}^{\mathrm{T}}X^{\mathrm{T}} + \bar{\varepsilon}\bar{\varepsilon}^{\mathrm{T}}) = X^{+}(-\bar{\varepsilon}\bar{\varepsilon}^{\mathrm{T}}X^{+\mathrm{T}}X^{\mathrm{T}} + \bar{\varepsilon}\bar{\varepsilon}^{\mathrm{T}}) = 0,$$

where, in the second equality, we have made use of the identity $XX^{+}\bar{\varepsilon} = \bar{\varepsilon}$, which, in turn, is due to the fact that $\bar{\varepsilon} \in \mathcal{R}(X^{+\mathrm{T}}X^{+}) = \mathcal{R}((XX^{\mathrm{T}})^{+}) = \mathcal{R}(XX^{\mathrm{T}}) = \mathcal{R}(X)$ and XX^{+} is the projection onto $R(X)$. This shows that (2.3) holds and hence $0 \in \partial\bar{F}(\bar{C})$.

Remark 1. Proposition 3 shows that the linear minimax estimator coincides with the ordinary least squares (OLS-) estimator of β, at least in situations where $\lambda_{\min}(X^{\mathrm{T}}X) \geq 1$. The latter condition will typically be fulfilled in reasonable models when the number n of observations is large, since $\lambda_{\min}(X^{\mathrm{T}}X) \to \infty$ as $n \to \infty$ is a necessary condition for weak consistency of the OLS–estimate (cf. Drygas [2]).

Remark 2. As in Christopeit and Helmes [1], one may elaborate further (2.3), using CarathCodory's theorem, to obtain sort of a *spectral equation*, which seems, however, not to allow a closed form solution, except for the special case considered in Proposition 3.

Remark 3. Let Π denote the set of all probability measures on T and consider the function

$$H(C, \pi) = \int_T \|\beta - Cy\|^2\pi\mathrm{d}\beta, \mathrm{d}\varepsilon) = \int_T F(C; \beta, \varepsilon)\pi(\mathrm{d}\beta, \mathrm{d}\varepsilon), \quad C \in I\!R^{K\times n}, \pi \in \Pi.$$

Apparently,

$$\sup_{\pi\in\Pi} H(C, \pi) = \max_{\|\beta\|,\|\varepsilon\|\leq 1} F(C; \beta, \varepsilon) = \bar{F}(C),$$

where the sup is adopted at any $\pi \in \Pi(C)$. Hence any solution of the minimax problem (2.1) also provides a solution to the problem

$$\min_{C \in \mathbb{R}^{K \times n}} \max_{\pi \in \Pi} H(C, \pi) = \max_{\pi \in \Pi} H(\bar{C}, \pi). \tag{2.5}$$

(2.5) may be thought of as a Bayesian version of (2.1), with the max taken with respect to all prior distributions on the (β, ε) parameter space. Let us note in passing that for (2.5) it can be shown that

$$\max_{\pi \in \Pi} \min_{C \in \mathbb{R}^{K \times n}} H(C, \pi) = \min_{C \in \mathbb{R}^{K \times n}} \max_{\pi \in \Pi} H(C, \pi)$$

(cf. Rockafellar [4]), where as for (2.1) minimax will generally be different from maximin.

3. JOINT RESTRICTIONS FOR PARAMETER AND DISTURBANCES

We consider again the linear regression model (1.1) $y = X\beta + \epsilon$. The a priori constants are now supposed to be of the form

$$\|\varepsilon\|^2 + \gamma \|\beta\|^2 \leq 1$$

for some positive known constant γ. Our objetive is to determine a linear estimator $\hat{\beta} = \hat{C} y$ minimizing the loss function

$$\varphi(C) = \max_{\|\epsilon\|^2 + \gamma \|\beta\|^2 \leq 1} (Cy - \beta)^{\mathrm{T}} V (Cy - \beta),$$

where V is a positive definite symmetric $L \times K$ matrix. The vector $\hat{\beta} = \hat{C} y$ is called the minimax estimator for the vector β.

As in Girko [7] we obtain that

$$\begin{aligned} &\min_{C \in K^{K \times n}} \max_{\|\epsilon\|^2 + \gamma \|\beta\|^2 \leq 1} (Cy - \beta)^{\mathrm{T}} V (Cy - \beta) \\ &\quad = \lambda_{\max} \Big\{ \gamma^{-1} \sqrt{V} (\hat{C} X - I)(\hat{C} X - I)^{\mathrm{T}} \sqrt{V} + \sqrt{V} \hat{C} \hat{C}^{\mathrm{T}} \sqrt{V} \Big\}, \end{aligned}$$

and that the matrix $\hat{C}$ satisfies the spectral equation

$$\Big\{ X (\hat{C} X - I)^{\mathrm{T}} \gamma^{-1} + \hat{C}^{\mathrm{T}} \Big\} \sum_{k=1}^{j} \sqrt{V} \vec{\nu}_k \vec{\nu}_k^{\mathrm{T}} \sqrt{V} p_k = 0, \tag{3.1}$$

where $\vec{\nu}_k,\ k = 1, \ldots, j$ are the orthogonal eigenvectors corresponding to the maximal j-multiple eigenvalue of the matrix

$$\gamma^{-1} \sqrt{V} (\hat{C} X - I)(\hat{C} X - I)^{\mathrm{T}} \sqrt{V} + \sqrt{V} \hat{C} \hat{C}^{\mathrm{T}} \sqrt{V},\ p_k > 0,\ \sum_{k=1}^{j} p_k = 1.$$

One of the solution of equation (3.1) is

$$\hat{C} = X^{\mathrm{T}} (\gamma I + X X^{\mathrm{T}})^{-1}.$$

4. FILTER PROBLEMS: STATIC CASE

In this section we develop an approach to the estimation of a solution of a system of equations with indefinite coefficients and random errors in a system of observations.

Let the system of linear equations

$$Ax = h + \xi_1,$$

be given, where A is a regular $m \times m$ matrix, x, h, ξ_1 are the vectors of dimension m, the matrix A and the vector h are known; while ξ_1 is an unknown vector of internal disturbances.

Assume that the observed vector y of dimension n is connected with the vector x by the equation

$$y = \Xi x + \xi_2,$$

where Ξ is known $(n \times m)$ matrix, ξ_2 is an unknown vector of dimension n, and ξ_1 and ξ_2 satisfy the inequality

$$\|\xi_1\|^2 + \|\xi_2\|^2 \leq 1.$$

The problem is to estimate x (optimally in a certain sense) by a linear transform of y. More precisely, we seek for an $m \times n$ matrix $\hat{K}$ and a vector $\hat{l}$ of dimension m which minimize the loss function

$$\varphi(K, l) = \max_{\|\xi_1\|^2 + \|\xi_2\|^2 \leq 1} \|x - Ky - l\|^2.$$

The vector $\hat{x} = \hat{K}y + \hat{l}$ is called a spectral or *minimax estimator* of the vector x.

Without loss of generality the vector h can be chosen to be zero.

PROPOSITION 4. *If the matrix A is regular, then*

$$\begin{aligned} &\min_{\substack{K \in R^{m \times n},\\ l \in R^m}} \max_{\|\xi_1\|^2 + \|\xi_2\|^2 \leq 1} \|x - Ky - l^*\|^2 \\ &\qquad = \lambda_{\max}\big\{(I - \hat{K}\Xi)A^{-1}{A^{-1}}^{\mathrm{T}}(I - \hat{K}\Xi)^{\mathrm{T}} + \hat{K}\hat{K}^{\mathrm{T}}\big\} \end{aligned} \tag{4.1}$$

and the matrix $\hat{K}$ satisfies the equation

$$\big\{-\Xi A^{-1}A^{-1\mathrm{T}}(I - \hat{K}\Xi)^{\mathrm{T}} + \hat{K}\big\} \sum_{k=1}^{s} p_k \varphi_k(D) \phi_k^{\mathrm{T}}(D) = 0, \tag{4.2}$$

where φ_k are orthonormal eigenvectors corresponding to the s-multiple maximal eigenvalue of the matrix $D = (I - \hat{K}\Xi)A^{-1}A^{-1\mathrm{T}}(I - \hat{K}\Xi)^{\mathrm{T}} + \hat{K}\hat{K}^{\mathrm{T}}$ and $\hat{l} = 0$, $p_k > 0$, $\sum_{k=1}^{s} p_k = 1$.

The proof is almost the same as in Girko [7]. It is sufficient to note that, as there, we get

$$\max_{\|\xi_1\|^2 + \|\xi_2\|^2 \leq 1} \|x - Ky\|^2 = \lambda_{\max}\big\{(I - K\Xi)A^{-1}A^{-1\mathrm{T}}(I - K\Xi)^{\mathrm{T}} + KK^{\mathrm{T}}\big\}.$$

5. FILTER PROBLEMS: THE DYMAMIC RECURSIVE CASE

In this section we consider the problem of estimation the state of a dynamical systems evolving in discrete time according to the difference equations

$$\begin{aligned} x_{i+1} &= x_i + A_i x_i + \xi_i, \\ y_i &= C_i x_i + \eta_i; \quad i = 1, 2, \ldots \end{aligned} \tag{5.1}$$

where x_i are m-dimensional unobservable state vectors; y_i are n-dimensional vectors of observed variables; A_i and C_i are matrices of dimension $m \times m$ and $n \times m$; ξ_i and η_i are error vectors of dimensions m and n, respectively, which satisfy the inequality

$$(\xi, \eta) \in G, \; G = \left\{ (\xi, \eta) : \; \sum_{i=1}^{k} \left(\|\xi_i\|^2 + \|\eta_i\|^2 \right) \leq 1 \right\} \text{ for some fixed } k$$

(where we have put $\xi = (\xi_1, \ldots, \xi_k)^{\mathrm{T}}$, and similarly η).

Consider the following problem of estimating the state x_k: to find matrices $\hat{K}_i$ of dimension $m \times n$ and a vector l of dimension m which minimize the expression

$$\varphi(K_1, \ldots, K_k; \, l) = \max_{(\xi,\eta) \in G} \left\| x_{k+1} - \sum_{i=1}^{k} K_i y_i - l \right\|. \tag{5.2}$$

The vector $\hat{x}_{k+1} = \sum_{i=1}^{k} \hat{K}_i y_i + \hat{l}$ is called linear minimax estimator of x_{k+1}.

Proposition 5. *Under the above formulated assumptions*

$$\min_{\substack{K_i \in R^{m \times n}; \\ l \in R^m}} \; \max_{(\xi,\eta) \in G} \left\| x_{k+1} - \sum_{i=0}^{k} K_i y_i - l \right\|^2 = \lambda_{\max} \left\{ \sum_{i=1}^{k} (Z_{i+1} Z_{i+1}^{\mathrm{T}} + \hat{K}_i \hat{K}_i^{\mathrm{T}}) \right\}, \tag{5.3}$$

where the matrices Z_i satisfy the recursive equations

$$Z_{p+1}(I + A_p) = Z_p + \hat{K}_p C_p; \; p = 1, \ldots, k; \; Z_{k+1} = I.$$

Moreover, $\hat{l} = Z_1 x_1$ and the matrices $\hat{K}_i, \; i = 1, \ldots, k$ satisfy the system of equations

$$\sum_{l=1}^{s} p_l \left[\hat{K}_p^{\mathrm{T}} + C_p S_p \right] \varphi_l \varphi_l^{\mathrm{T}} = 0; \; p = 1, \ldots, k \tag{5.4}$$

where $p_l > 0$; $\sum_{j=0}^{s} p_l = 1$, φ_k, $k = 1, \ldots, s$ are orthonormal eigenvectors which correspond to the maximal s-multiple eigenvalue $\lambda_{\max}$ of the matrix $\sum_{i=1}^{k} [Z_{i+1} Z_{i+1}^{\mathrm{T}} + \hat{K}_i \hat{K}_i^{\mathrm{T}}]$, and the matrices S_p satisfying the system of equations

$$S_{p+1} = S_p + A_p S_p - Z_{p+1}^{\mathrm{T}}; \; p = 1, \ldots, k; \; S_1 = 0, \tag{5.5}$$

one of the solution of the equation (5.4) is

$$\hat{K}_p^{\mathrm{T}} = -C_p S_p, \; p = 1, \ldots, k.$$

Proof. It is obvious that

$$\left\|x_{k+1}-\sum_{i=1}^{k} K_i y_i - l\right\|^2 = \left\|x_{k+1}-\sum_{i=1}^{k} K_i C_i x_i - \sum_{i=0}^{k} K_i \eta_i - l\right\|^2.$$

Consider the system of recursive equations

$$Z_{p+1} = Z_p - Z_{p+1}A_p + K_p C_p;\ p = 1,\dots,k$$

with the initial condition $Z_{k+1} = I$. Then, using (5.1), after obvious transformations we have

$$\begin{aligned} x_{k+1}-\sum_{i=1}^{k} K_i C_i x_i &= x_{k+1} - \sum_{i=1}^{k}(Z_{i+1}-Z_i)x_i - \sum_{i=1}^{k} Z_{i+1}A_i x_i \\ &= x_{k+1} - \sum_{i=1}^{k} Z_{i+1}(x_{i+1}-A_i x_i - \xi_i) - \sum_{i=1}^{k} Z_{i+1}A_i x_i + \sum_{i=1}^{k} Z_i x_i \\ &= Z_1 x_1 + \sum_{i=1}^{k} Z_{i+1}\xi_i. \end{aligned}$$

Therefore

$$\left\|x_{k+1}-\sum_{i=1}^{k} K_i y_i - l\right\|^2 = \left\|Z_1 x_1 + \sum_{i=1}^{k} Z_{i+1}\xi_i + \sum_{i=1}^{k} K_i \eta_i - l\right\|^2.$$

As in Girko [7] we get

$$\begin{aligned} &\min_{\substack{K_i\in R^{m\times n};\\ l\in R^m}} \max_{(\xi,\eta)\in G} \left\|x_{k+1}-\sum_{i=1}^{k} K_i y_i - l\right\|^2 \\ &\qquad = \min_{K_i\in R^{m\times n}} \lambda_{\max}\left\{\sum_{i=1}^{k}(Z_{i+1}Z_{i+1}^{\mathrm{T}} + \hat{K}_i\hat{K}_i^{\mathrm{T}})\right\},\ \hat{l} = Z_1\vec{x}_1, \end{aligned}$$

and unknown matrices $\hat{K}_i$ satisfy the equation

$$\sum_{q=1}^{s} \vec{\varphi}_q^{\mathrm{T}} p_q \sum_{i=1}^{k}(\tilde{Z}_{i+1}Z_{i+1}^{\mathrm{T}} + \Theta_i\hat{K}_i^{\mathrm{T}})\varphi_q = 0, \tag{5.6}$$

where $\varphi_k,\ k = 1,\dots,s$ are orthonormal eigenvectors which correspond to the maximal s-multiple eigenvalue $\lambda_{\max}$ of the matrix $\sum_{i=1}^{k}[Z_{i+1}Z_{i+1}^{\mathrm{T}} + \hat{K}_i\hat{K}_i^{\mathrm{T}}]$, Θ_i are arbitrary matrices which have the same dimension as matrices K_i, and the matrices $\tilde{Z}_{i+1}$ satisfy the equations

$$\tilde{Z}_{i+1} = \tilde{Z}_i - \tilde{Z}_{i+1}A_i + \Theta_i C_i;\quad \tilde{Z}_{k+1} = 0;\ i = 1,\dots,k.$$

Obviously

$$
\begin{aligned}
\sum_{i=1}^{k} \tilde{Z}_{i+1} Z_{i+1}^{\mathrm{T}} &= \sum_{i=1}^{k} (\tilde{Z}_{i+1} Z_{i+1}^{\mathrm{T}} + \tilde{Z}_{i+1} S_{i+1} - \tilde{Z}_i S_i) \\
&= \sum_{i=1}^{k} \big[\tilde{Z}_{i+1} Z_{i+1}^{\mathrm{T}} + \tilde{Z}_{i+1}(S_i + A_i S_i - Z_{i+1}^{\mathrm{T}}) - \tilde{Z}_i S_i\big] \\
&= \sum_{i=1}^{k} \big[(\tilde{Z}_i - \tilde{Z}_{i+1} A_i + \Theta_i C_i) S_i + \tilde{Z}_{i+1} A_i S_i - \tilde{Z}_i S_i)\big] \\
&= \sum_{i=1}^{k} \Theta_i C_i S_i .
\end{aligned}
$$

Using this equality and the auxiliary systems of equations (5.5) we obtain that (5.6) equals

$$
\sum_{q=1}^{s} \vec{\varphi}_q^{\mathrm{T}} p_q \sum_{i=1}^{k} \Theta_i (\hat{K}_i^{\mathrm{T}} + C_i S_i) \varphi_q = 0.
$$

From this equation we obtain all assertions of Proposition 5.

6. FILTER PROBLEMS: THE DYNAMIC CASE

Let the system of equations

$$
\frac{\mathrm{d}x(t)}{\mathrm{d}t} = A(t)x(t) + \xi_1(t), \quad x(0) = a \tag{6.1}
$$

be given, where $A(t)$ is a square $n \times n$ matrix, whose entries are piecewise continuous function, $x(t)$ is a vector of dimension n of states of this system, $\xi_1(t)$ is the vector of perturbations of dimension n and the vector a is given.

Let the vector of observations $y(t),\ 0 \le t \le T$ of dimension m be observed and let it be related with vector $x(t)$:

$$
y(t) = C(t)x(t) + \xi_2(t), \quad 0 \le t \le T, \tag{6.2}
$$

where $C(t)$ is an $m \times n$ matrix, whose entries are piecewise continuous functions, $T \ge 0$ is some constant, $\xi_2(t)$ is a vector of perturbations of dimension m.

Assume that the components of vectors $\xi_1(t)$, $\xi_2(t)$ are piecewise continuous functions and that the vectors $\xi_1(t)$ and $\xi_2(t)$ belong to the domain

$$
G = \Big\{ \xi_1(\cdot), \xi_2(\cdot) : \int_0^T \big[\|\xi_1(t)\|^2 + \|\xi_2(t)\|^2\big] \mathrm{d}t \le 1 \Big\},
$$

where $\|\xi_1(t)\|^2 = \xi_1^{\mathrm{T}}(t)\xi_1(t)$.

Let R and Q be the sets of matrices $K(t)$ of dimension $m \times n$ and vectors $l(T)$ of dimension n and the entries of the matrix $K(t)$ belong to the Hilbert space $L_2[0,T]$.

The problem of estimating the state $x(T)$ is on finding matrices $\hat{K}(u)$ and a vector $\hat{l}(T)$ such that

$$\begin{aligned}\min_{\substack{K(u)\in R,\\ l(T)\in Q}} \max_{\xi_1(\cdot),\xi_2(\cdot)\in G} & \left\| x(T) - \int_0^T K(u)y(u)\,\mathrm{d}u - \hat{l}(T) \right\|^2 \\ & = \left\| x(T) - \int_0^T \hat{K}(u)y(u)\,\mathrm{d}u - \hat{l}(T) \right\|^2 .\end{aligned} \tag{6.3}$$

The expression

$$x(T) = \int_0^T \hat{K}(u)y(u)\,\mathrm{d}u - \hat{l}(T), \tag{6.4}$$

where $\hat{K}(u)$ and $\hat{l}(T)$ minimize the expression above, is called the spectral or minimax estimator of the state $x(T)$.

Proposition 6. *Under the above assumptions, the solutions $\hat{K}(u)$, $\hat{l}(T)$ of the equation (6.3) can be found from the equations:*

$$[K^{\mathrm{T}}(u) + C(u)S(u)] \sum_{k=1}^{s} p_k \varphi_k \varphi_k^{\mathrm{T}} = 0, \quad 0 \le u \le T, \tag{6.5}$$

$$\frac{\mathrm{d}S(t)}{\mathrm{d}t} = A(t)S(t) - Z^{\mathrm{T}}(t), \ S(0) = 0, \ \hat{l}(T) = Z(0)a, \tag{6.6}$$

$$\frac{\mathrm{d}Z(t)}{\mathrm{d}t} = -Z(t)A(t) + \hat{K}(t)C(t), \ Z(t) = 1, \tag{6.7}$$

where $\sum_{k=1}^{s} p_k = 1$, $p_k > 0$, φ_k, $k = 1, \ldots, s$ are the orthonormal eigenvectors corresponding to the maximal s-multiple eigenvalue of matrix

$$\int_0^{\mathrm{T}} \left[Z(t)Z(t)^{\mathrm{T}} + K(t)K(t)^{\mathrm{T}}\right] \mathrm{d}t.$$

Proof. It is obvious that

$$\begin{aligned} & \left\| x(T) - \int_0^T K(u)y(u)\,\mathrm{d}u - l(T) \right\| \\ & \quad = \left\| x(T) - \int_0^T K(t)C(t)x(t)\,\mathrm{d}t - \int_0^T K(t)\xi_2(t)\,\mathrm{d}t - l(T) \right\| .\end{aligned} \tag{6.8}$$

Using the system of differential equations (6.1)–(6.7) we have

$$\begin{aligned} x(T) = Z(T)x(T) &= \int_0^T \frac{\mathrm{d}[Z(t)x(t)]}{\mathrm{d}t}\mathrm{d}t + Z(0)a \\ &= \int_0^T K(t)C(t)\,x(t)\,\mathrm{d}t + \int_0^T Z(t)\xi_1(t)\,\mathrm{d}t + Z(0)a. \end{aligned}$$

From this equality and (6.8) we get

$$\left\| x(T) - \int_0^T \hat{K}(u)y(u)\,\mathrm{d}u - \hat{l}(T) \right\|^2 = \left\| Z(0)a - \hat{l}(T) + \int_0^T [Z(t)\xi_1(t) - K(t)\xi_2(t)]\,\mathrm{d}t \right\|^2.$$

Hence, using the Rayleigh formula we find that $\hat{l}(T) = Z(0)a$,

$$\begin{aligned}&\min_{\substack{K(u)\in R,\\ l(T)\in Q}} \max_{\xi_1(\cdot),\xi_2(\cdot)\in G} \left\| x(T) - \int_0^T K(u)y(u)\,\mathrm{d}u - \hat{l}(T) \right\|^2 \\ &\qquad = \min_{\substack{K(u)\in R,\\ l(T)\in Q}} \lambda_{\max} \left[\int_0^T \left[Z(t)Z(t)^{\mathrm{T}} + K(t)K(t)^{\mathrm{T}} \right] \mathrm{d}t \right].\end{aligned}$$

We see that the unknown matrix $K(s)$ satisfies the equality

$$\operatorname{Tr} \int_0^T \left[\tilde{Z}(t)Z(t)^{\mathrm{T}} + \Theta(t)K(t)^{\mathrm{T}} \right] \mathrm{d}t \sum_{k=1}^{s} p_k \varphi_k \varphi_k^{\mathrm{T}} = 0,$$

where $\Theta(t)$ is an arbitrary matrix from the set R and the matrix $\tilde{Z}(t)$ satisfies the equality

$$\frac{\mathrm{d}\tilde{Z}(T)}{\mathrm{d}t} = -\tilde{Z}(t)A(t) + \Theta(t)C(t), \quad \tilde{Z}(T) = 0.$$

Using the auxiliary system of equations

$$\frac{\mathrm{d}S(t)}{\mathrm{d}t} = A(t)S(t) - Z^{\mathrm{T}}(t), \quad S(0) = 0,$$

we get

$$\int_0^{\mathrm{T}} \tilde{Z}(t)Z^{\mathrm{T}}(t)\,\mathrm{d}t = \int_0^T \Theta(t)C(t)S(t)\,\mathrm{d}t.$$

Therefore, for the matrix $K(s)$, we have the inequality

$$\operatorname{Tr} \left\{ \int \theta(t) \left[C(t)S(t) + K^{\mathrm{T}} \right] \mathrm{d}t \sum_{k=1}^{s} p_k \varphi_k \varphi_k^{\mathrm{T}} \right\} = 0.$$

From this equation in virtue of arbitrariness of matrix Θ the Proposition 6 follows.

7. ESTIMATION OF LINEAR REGRESSION MODELS IN HILBERT SPACE

Denote the real separable Hilbert spaces of elements by $H_1,\ H_2$. Let B_1 be the Banach space of operators, which maps H_2 into H_1, let B_2 be the Banach space of linear real operators which maps H_1 into H_1, and B_3 is the Banach space of linear real operators which maps H_2 into H_2.

Assume that the linear model of regression $y = Xc + \varepsilon$ in the Hilbert space H_2 is given, where c is an unknown element in the Hilbert space H_2, y is an element from the Hilbert space H_1, X is a linear operator which maps H_2 into H_1, ε is a element of unobservable perturbations from Hilbert space H_1. Let the elements c and ε satisfy the inequalities

$$(c, c) + (\varepsilon, \varepsilon) \leq a,$$

where $0 < a < \infty$. By means of a linear transformation of the element y: $Ty + t$ we find an operator $\hat{T} \in B_1$ and an element $\hat{t} \in H_1$ such that the expression (loss function)

$$f(T, t) = \sup_{(c,c)+(\varepsilon,\varepsilon)\leq a} ([Ty + t - c], V[Ty + t - c]),$$

where V is a symmetric nonnegative definite operator from B_3, will take a minimal value. The element $\hat{c} = Ty + \hat{t}$ is called the S-estimator of the element c.

As in Girko [7] we prove the following statements.

PROPOSITION 7. *If V, X, D^{-1} are operators of trace class, then*

$$\inf_{\substack{T\in B_1,\\ t\in H_2}} f(T, t) = \alpha\lambda_1 \big[\sqrt{V}(I - \hat{T}_i X)(I - \hat{T}_i X)^{\mathrm{T}} \sqrt{V} + \sqrt{V} \hat{T}_i \hat{T}_i^{\mathrm{T}} \sqrt{V}\big],$$

where $V^{1,2}\hat{t} = 0$, λ_1 is maximal eigenvalue of multiplicity i, I is identity operator and the operators T_i satisfy equation

$$\big[X(\hat{T}_i X - I)^{\mathrm{T}} + T^{\mathrm{T}}\big] a \sum_{k=1}^{i} \sqrt{V} E_k \sqrt{V} p_k = 0,$$

here $p_k > 0$, $\sum_{k=1}^{i} p_k = 1$, E_k, $k = 1, \ldots, i$ are orthogonal projectors which correspond to the eigenvalue λ_1.

REFERENCES

1. Christopeit and Helmes (1991). On minimax estimation in linear regression models with elipsoidal constraints. University of Bonn, Discussion Paper, SFB 303, Econometrics Unit.
2. Drygas, H. (1976). Weak and strong consistency of least squares estimators for families of linear regressions. Z. f. Wahrscheinlichkeitstheorie u. Verw. Gebiete 34, 119–127.
3. Joffe & Levin (1974). Subdifferentials of convex functions. Transactions of the Moscow Mathematical Society 26.
4. Rockefellar, R. T. (1974). Conjugate duality and optimization, SIAM, Philadelphia.
5. H. Drygas. Minimax Linear Estimation in a convex linear model. – A paradoxon in Decision Theory? *Statistics and Decision*, **10**, 63–66 (1992).
6. H. Drygas. On an Extension of the Girko Equality in Linear Minimax Estimation, *Proceedings of International Conference Probastat '91 on Probability and Mathematical Statistics*, (Ptzman, Volaufovt, eds.), Bratislava/Slovakia, (1992).
7. V. L. Girko. On a class of spectral estimators, *Theory of Probability and mathematical statistics* **47**, 15–22 (1993).
8. V. L. Girko. Spectral Theory of Estimation, Proceedings of the Oldenburg Workshop on Minimax Estimation. To appear in: *Acta Applicandae Mathematicae* (1995).
9. A. G. Nakonechny. To minimax theory of estimation of functionals of solutions of operator equation, *Rep. Ukr. Acad. Sci. Ser. Math. Mech.* **5**, *69–72 (1982). (in Russian).*