Kybernetika

Andrej Pázman

Distribution of the weighted L.S. estimates in nonlinear models with symmetrical errors

Kybernetika, Vol. 24 (1988), No. 6, 413--427

Persistent URL: http://dml.cz/dmlcz/124188

KYBERNETIKA — VOLUME 24 (1988), NUMBER 6

DISTRIBUTION OF THE WEIGHTED L.S. ESTIMATES IN NONLINEAR MODELS WITH SYMMETRICAL ERRORS

ANDREJ PÁZMAN

The nonlinear regression model $\mathbf{y} = \boldsymbol{\eta}(\boldsymbol{\theta}) + \boldsymbol{\varepsilon}$ with the error vector $\boldsymbol{\varepsilon}$ having an elliptically symmetrical probability distribution is considered. An approximative formula for the non-asymptotical (= small sample) probability density of the weighted L. S. estimates of $\boldsymbol{\theta}$ is obtained by geometrical methods. The considered weights are general (i.e. not related to the variance matrix $\boldsymbol{\Sigma}$ of $\boldsymbol{\varepsilon}$). The difference between the true and the approximative densities is evaluated. Earlier author's results are thus extended from the case of normal errors, and of weights depending on $\boldsymbol{\Sigma}$, to a more general case.

1. INTRODUCTION

Let

$$\mathbf{y} = \boldsymbol{\eta}(\boldsymbol{\theta}) + \boldsymbol{\varepsilon} \tag{1}$$

be a nonlinear regression model. Here $\mathbf{y} := (y_1, \ldots, y_N)^T$ is the vector of the observed data, $\boldsymbol{\theta} := (\theta_1, \ldots, \theta_m)^T$ is the vector of unknown parameters, $m < N$, $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ where $\boldsymbol{\Theta}$ is the (given) parameter space which is an open subset of $\mathbb{R}^m$. The mapping $\boldsymbol{\eta}: \boldsymbol{\theta} \in \bar{\boldsymbol{\Theta}} \mapsto \boldsymbol{\eta}(\boldsymbol{\theta}) \in \mathbb{R}^N$, defined and finite on the closure $\bar{\boldsymbol{\Theta}}$ of the set $\boldsymbol{\Theta}$, is supposed to be known, continuous, and to have continuous second order derivatives on $\boldsymbol{\Theta}$. The vectors of the first order derivatives $\partial\boldsymbol{\eta}(\boldsymbol{\theta})/\partial\theta_1, \ldots, \partial\boldsymbol{\eta}(\boldsymbol{\theta})/\partial\theta_m$ are supposed to be linearly independent for every $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ (i.e. the model is regular).

In this paper we consider the case when the probability density of the error vector $\boldsymbol{\varepsilon}$ is elliptically symmetrical, with a zero mean $\mathsf{E}(\boldsymbol{\varepsilon}) = 0$, and a positive definite variance matrix $\boldsymbol{\Sigma}$, $\mathsf{Var}(\boldsymbol{\varepsilon}) = \boldsymbol{\Sigma}$, defining the elliptical symmetry. Such a probability density (with respect to the Lebesgue measure in $\mathbb{R}^N$) is given by the formula (cf. [5])

$$f(\boldsymbol{\varepsilon}) := \det^{-1/2}(\boldsymbol{\Sigma})\, h(\boldsymbol{\varepsilon}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon}) \tag{2}$$

where $h: \langle 0, \infty) \mapsto \langle 0, \infty)$ is a function such that

$$\int_0^\infty z^{N/2}\, h(z)\, \mathrm{d}z < \infty\,.$$

To ensure that $f(\boldsymbol{\varepsilon})$ is a probability density and that $\operatorname{Var}(\boldsymbol{\varepsilon}) = \boldsymbol{\Sigma}$ we have to suppose that

$$\int_{\boldsymbol{R}^N} h(\|\mathbf{v}\|^2)\,\mathrm{d}\mathbf{v} = 1$$

$$\int_{\boldsymbol{R}^N} h(\|\mathbf{v}\|^2)\,\|\mathbf{v}\|^2\,\mathrm{d}\mathbf{v} = N\,.$$

If the function h does not satisfy these two norming conditions, we can always find two positive numbers α and β such that the function $\mathbf{z} \mapsto \alpha h(\beta \mathbf{z})$ has the required properties. (We note that, like in Section 2, these two N-dimensional integrals can be changed to two onedimensional integrals when using spherical coordinates in $\mathbf{R}^N$.)

The set $\{\boldsymbol{\varepsilon}: f(\boldsymbol{\varepsilon}) = \text{const}\}$ is an ellipsoid in $\mathbb{R}^N$, therefore we speak about the elliptical symmetry. In the case of $\boldsymbol{\Sigma} = \mathbf{I}$, $f(\boldsymbol{\varepsilon})$ is spherically symmetrical. Another equivalent definition of the spherical symmetry is that $f(\boldsymbol{\varepsilon}) = f(\mathbf{U}\boldsymbol{\varepsilon})$ for every orthogonal $m \times m$ matrix $\mathbf{U}$ (i.e. such that $\mathbf{U}^{\mathrm{T}}\mathbf{U} = \mathbf{I}$). Thus spherically symmetrical densities are invariant to every rotation of the sample space of $\boldsymbol{\varepsilon}$.

Elliptically symmetrical distributions are studied in several papers [2, 5, 6], and we resume their properties in Section 2.

A special case of an elliptically symmetrical density is the normal density $N(0, \boldsymbol{\Sigma})$ with

$$h(t) = (2\pi)^{-N/2} \exp\{-t/2\}\,.$$

Other choices of the function $h(\cdot)$ are presented in Section 2.

A standard estimator of the vector $\boldsymbol{\theta}$ is the weighted least squares (= L. S.) estimator given by

(3) $$\hat{\boldsymbol{\theta}} := \hat{\boldsymbol{\theta}}(\mathbf{y}) := \arg\min\,[\mathbf{y} - \boldsymbol{\eta}(\boldsymbol{\theta})]^{\mathrm{T}}\,\mathbf{V}^{-1}[\mathbf{y} - \boldsymbol{\eta}(\boldsymbol{\theta})]\,,$$

where $\boldsymbol{\theta} \in \overline{\boldsymbol{\Theta}}$ and $\mathbf{V}$ is some given positive definite (= p.d.) matrix. Usually (if possible) the matrix $\mathbf{V}$ is proportional to the covariance matrix $\boldsymbol{\Sigma}$. This leads to an optimal unbiased estimator of $\boldsymbol{\theta}$ when the model (1) is linear (i.e. $\boldsymbol{\eta}(\boldsymbol{\theta}) = \mathbf{A}\boldsymbol{\theta} + \boldsymbol{a}$) (cf. [6]), and such a $\mathbf{V}$ is considered as preferable also in the nonlinear case. However, if $\boldsymbol{\Sigma}$ is unknown, the matrix $\mathbf{V}$ is to be chosen and hoc. Since the estimate (3) is not influenced by setting a matrix $c\mathbf{V}$ $(c > 0)$ instead of $\mathbf{V}$, we can always choose $\mathbf{V}$ such that it dominates the matrix $\boldsymbol{\Sigma}$, i.e. that

$$\boldsymbol{a}^{\mathrm{T}}\mathbf{V}^{-1}\boldsymbol{a} \leqq \boldsymbol{a}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{a}\,; \quad (\boldsymbol{a} \in \mathbb{R}^N)$$

(see Proposition 3).

The normal equations corresponding to (3) are

$$\frac{\partial[\mathbf{y} - \boldsymbol{\eta}(\boldsymbol{\theta})]^{\mathrm{T}}\,\mathbf{V}^{-1}[\mathbf{y} - \boldsymbol{\eta}(\boldsymbol{\theta})]}{\partial \theta_i} = 0\,; \quad (i = 1, \ldots, m)\,,$$

hence, if $\hat{\boldsymbol{\theta}}(\mathbf{y}) \in \boldsymbol{\Theta}$, it is a solution of

(4) $$[\mathbf{y} - \boldsymbol{\eta}(\boldsymbol{\theta})]^{\mathrm{T}}\,\mathbf{V}^{-1}\,\frac{\partial \boldsymbol{\eta}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{\mathrm{T}}} = \mathbf{0}\,.$$

In this paper we present an approximative nonasymptotical probability density of $\hat{\boldsymbol{\theta}}$, and we present a formula for the upper bound for the difference between the true and the approximative densities. Earlier author's results [7, 8] are thus extended from the case of normal errors to the case of elliptically symmetrical errors, and from the case of $\mathbf{V} = \boldsymbol{\Sigma}$ to the case or arbitrary, p.d. matrices $\mathbf{V}$ and $\boldsymbol{\Sigma}$. However, the main geometrical ideas remain unchanged since the elliptical symmetry has been important also in the investigation presented in [7, 8].

The approximative nonasymptotical probability density of $\hat{\boldsymbol{\theta}}$ proposed in this paper is equal to

$$q(\hat{\boldsymbol{\theta}} \mid \bar{\boldsymbol{\theta}}) := \frac{\det \mathbf{Q}(\hat{\boldsymbol{\theta}}, \bar{\boldsymbol{\theta}})}{\det^{1/12} \mathbf{B}(\hat{\boldsymbol{\theta}})} h_m(\|\mathbf{P}^{\hat{\boldsymbol{\theta}}}[\boldsymbol{\eta}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\eta}]\|_{\boldsymbol{\Sigma}}^2) \tag{5}$$

where

$$\boldsymbol{\eta} := \boldsymbol{\eta}(\bar{\boldsymbol{\theta}})$$

is the true mean of $\mathbf{y}$,

$$\mathbf{B}(\boldsymbol{\theta}) := \frac{\partial \boldsymbol{\eta}^{\mathrm{T}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{V}^{-1} \boldsymbol{\Sigma} \mathbf{V}^{-1} \frac{\partial \boldsymbol{\eta}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{\mathrm{T}}},$$

$$\mathbf{Q}(\boldsymbol{\theta}, \bar{\boldsymbol{\theta}}) := \mathbf{M}(\boldsymbol{\theta}) + [(\mathbf{I} - \mathbf{P}^{\boldsymbol{\theta}})(\boldsymbol{\eta}(\boldsymbol{\theta}) - \boldsymbol{\eta})]^{\mathrm{T}} \mathbf{V}^{-1} \frac{\partial^2 \boldsymbol{\eta}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\, \partial \boldsymbol{\theta}^{\mathrm{T}}} \tag{6}$$

$$\mathbf{M}(\boldsymbol{\theta}) := \frac{\partial \boldsymbol{\eta}^{\mathrm{T}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{V}^{-1} \frac{\partial \boldsymbol{\eta}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{\mathrm{T}}},$$

$$\mathbf{P}^{\boldsymbol{\theta}} := \boldsymbol{\Sigma} \mathbf{V}^{-1} \frac{\partial \boldsymbol{\eta}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{\mathrm{T}}} \mathbf{B}^{-1}(\boldsymbol{\theta}) \frac{\partial \boldsymbol{\eta}^{\mathrm{T}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{V}^{-1} \tag{7}$$

($\mathbf{P}^{\boldsymbol{\theta}}$ is a projector),

$$\|\boldsymbol{a}\|_{\boldsymbol{\Sigma}}^2 := \boldsymbol{a}^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \boldsymbol{a}\,; \quad (\boldsymbol{a} \in \mathbb{R}^N),$$

and where h_m: $\langle 0, \infty) \mapsto \langle 0, \infty)$ is defined by the formula

$$h_m(t) := \frac{\pi^{(N-m)/2}}{\Gamma\left(\dfrac{N-m}{2}\right)} \int_0^\infty u^{(N-m)/2-1}\, h(t+u)\, \mathrm{d}u\,. \tag{8}$$

The expression in (5) becomes simpler when $\boldsymbol{\Sigma} = \mathbf{V}$. Then $\mathbf{M}(\boldsymbol{\theta}) = \mathbf{B}(\boldsymbol{\theta}) =$ the Fisher information matrix for the case of normal errors, and $\mathbf{Q}(\boldsymbol{\theta}, \bar{\boldsymbol{\theta}})$ is the information matrix $\mathbf{M}(\boldsymbol{\theta})$ corrected by a term reflecting the curvature of the model (1). ($\mathbf{Q}(\hat{\boldsymbol{\theta}}, \bar{\boldsymbol{\theta}})$ is a measure of the observed information gained from the experiment when $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\mathbf{y})$ is obtained from the observation and $\bar{\boldsymbol{\theta}}$ is the true value of $\boldsymbol{\theta}$ (cf. [9]).)

In the case that the model (1) is linear, $\boldsymbol{\eta}(\boldsymbol{\theta}) = \mathbf{A}\boldsymbol{\theta}$, $q(\hat{\boldsymbol{\theta}} \mid \bar{\boldsymbol{\theta}})$ is equal to the exact probability density of $\hat{\boldsymbol{\theta}}$. In the case that $\mathbf{V} = \boldsymbol{\Sigma}$, it is equal to

$$q(\hat{\boldsymbol{\theta}} \mid \bar{\boldsymbol{\theta}}) = \det^{1/2}(\mathbf{M})\, h_m[(\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}})^{\mathrm{T}} \mathbf{M}(\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}})]\,,$$

where $\mathbf{M} := \mathbf{A}\boldsymbol{\Sigma}^{-1}\mathbf{A}^{\mathrm{T}}$ is the information matrix. In the normal case we obtain the well known formula

$$q(\hat{\boldsymbol{\theta}} \mid \bar{\boldsymbol{\theta}}) = (2\pi)^{m/2} \det^{1/2}(\mathbf{M}) \exp\{-\tfrac{1}{2}(\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}})^{\mathrm{T}} \mathbf{M}(\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}})\}\,.$$

In the general case the approximative density $q(\hat{\boldsymbol{\theta}} \mid \bar{\boldsymbol{\theta}})$ is invariant to the change of parameters $\boldsymbol{\beta} = \boldsymbol{\beta}(\boldsymbol{\theta})$, i.e.

$$q(\hat{\boldsymbol{\theta}} \mid \bar{\boldsymbol{\theta}}) = \left|\det\left(\frac{\partial \boldsymbol{\beta}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{\mathrm{T}}}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}\right)\right| q(\hat{\boldsymbol{\beta}} \mid \bar{\boldsymbol{\beta}})\,,$$

where $q(\hat{\boldsymbol{\beta}} \mid \bar{\boldsymbol{\beta}})$ is obtained by putting into the expression (5) the function $\boldsymbol{\nu}(\boldsymbol{\beta}) := := \boldsymbol{\eta}[\boldsymbol{\theta}^{-1}(\boldsymbol{\beta})]$ and its derivatives instead of the function $\boldsymbol{\eta}(\boldsymbol{\theta})$.

Example. (The contaminated normal nonlinear regression.)

Suppose that the probability density of $\boldsymbol{\varepsilon}$ is equal to

$$f(\boldsymbol{\varepsilon}) = (2\pi)^{-N/2}\left[(0{\cdot}9)\exp\{-\tfrac{1}{2}\|\boldsymbol{\varepsilon}\|^2\} + \frac{(0{\cdot}1)}{10^{N/2}}\exp\{-\tfrac{1}{20}\|\boldsymbol{\varepsilon}\|^2\}\right]$$

and consider the non-weighted L. S. estimates. Hence $\mathbf{V} = \boldsymbol{\Sigma} = \mathbf{I}$, and

$$h(t) = (2\pi)^{-N/2}[(0{\cdot}9)\exp\{-\tfrac{1}{2}t\} + 10^{-N/2-1}\exp\{-\tfrac{1}{20}t\}]\,.$$

Consequently

$$h_m(t) = (2\pi)^{-m/2}[(0{\cdot}9)\exp\{-\tfrac{1}{2}t\} + 10^{-m/2-1}\exp\{-\tfrac{1}{20}t\}]$$

because $h_m(\sum_{i=1}^{m} \varepsilon_i^2)$ is the m-dimensional marginal of $f(\boldsymbol{\varepsilon})$ (see Section 2). Further

$$\mathbf{P}^{\boldsymbol{\theta}} = \frac{\partial \boldsymbol{\eta}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{\mathrm{T}}} \mathbf{M}^{-1}(\boldsymbol{\theta}) \frac{\partial \boldsymbol{\eta}^{\mathrm{T}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\,,$$

$$\mathbf{M}(\boldsymbol{\theta}) = \mathbf{B}(\boldsymbol{\theta}) = \frac{\partial \boldsymbol{\eta}^{\mathrm{T}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \boldsymbol{\eta}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{\mathrm{T}}}\,,$$

and

$$q(\hat{\boldsymbol{\theta}} \mid \bar{\boldsymbol{\theta}}) = (2\pi)^{-m/2} \frac{\det\left[[\mathbf{M}(\hat{\boldsymbol{\theta}}) + [\boldsymbol{\eta}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\eta}]^{\mathrm{T}} (\mathbf{I} - \mathbf{P}^{\hat{\boldsymbol{\theta}}}) \dfrac{\partial^2 \boldsymbol{\eta}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\, \partial \boldsymbol{\theta}^{\mathrm{T}}}\right]}{\det^{1/2} \mathbf{M}(\hat{\boldsymbol{\theta}})} \times$$

$$\times\ [(0{\cdot}9)\exp\{-\tfrac{1}{2}\|\mathbf{P}^{\hat{\boldsymbol{\theta}}}[\boldsymbol{\eta}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\eta}]\|^2\} + 10^{-m/2-1}\exp\{-\tfrac{1}{20}\|\mathbf{P}^{\hat{\boldsymbol{\theta}}}[\boldsymbol{\eta}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\eta}]\|^2\}]$$

Computing point by point both components of $q(\hat{\boldsymbol{\theta}} \mid \bar{\boldsymbol{\theta}})$, we can evaluate the influence of the contamination on the least squares in a gaussian nonlinear model.

2. PROPERTIES OF ELLIPTICALLY (SPHERICALLY) SYMMETRICAL DENSITIES

We write: $\mathbf{y} \sim S_N(\boldsymbol{\eta}, \boldsymbol{\Sigma}, h)$ iff $\mathbf{y}$ has the density

$$f_Y(\mathbf{y}) = \det^{-1/2}(\boldsymbol{\Sigma})\, h[(\mathbf{y} - \boldsymbol{\eta})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \boldsymbol{\eta})]\,. \tag{9}$$

This density has all moments up to the kth order iff

(10) $$\int_0^\infty u^{\frac{N+k}{2}-1} h(u)\,\mathrm{d}u < \infty$$

(cf. [5]). If $k \geqq 1$, we have $\mathsf{E}(\mathbf{y}) = \boldsymbol{\eta}$. If $k \geqq 2$, we have $\mathsf{Var}(\mathbf{y}) = \boldsymbol{\Sigma}$. (See Section 1 for the norming conditions on h.)

If $\mathbf{z} = \mathbf{A}\mathbf{y}$, where $\mathbf{A}$ is an $N \times N$ nonsingular matrix, then

$$\mathbf{z} \sim S_N(\mathbf{A}\boldsymbol{\eta}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^{\mathrm{T}}, h)$$

(cf. [5]). Consequently, if $\mathbf{y} \sim S_N(\boldsymbol{\eta}, \boldsymbol{\Sigma}, h)$, then there is a matrix $\mathbf{A}$ such that $\mathbf{z} = \mathbf{A}(\mathbf{y} - \boldsymbol{\eta}) \sim S_N(\mathbf{0}, \mathbf{I}, h)$.

If $\mathbf{y} \sim S_N(\boldsymbol{\eta}, \boldsymbol{\Sigma}, h)$, then

$$\mathbf{y} = \boldsymbol{\eta} + l\boldsymbol{\Sigma}^{1/2}\boldsymbol{u},$$

where the vector $\boldsymbol{u}$ is uniformly distributed on the unit sphere $\{\mathbf{z}: \mathbf{z} \in \mathbb{R}^N, \|\mathbf{z}\| = 1\}$, and where l is a nonnegative random variable which is independent of $\boldsymbol{u}$ (cf. [6]).

If $\boldsymbol{\varepsilon} \sim S_N(\mathbf{0}, \mathbf{I}, h)$, then the marginal density of $(\varepsilon_{i_1}, \ldots, \varepsilon_{i_m})$ is equal to

$$h_m\Big(\sum_{k=1}^{m} \varepsilon_{i_k}^2\Big)$$

where

$$h_m(t) := \int_{\mathbb{R}^{N-m}} h(t + \|\mathbf{v}\|^2)\,\mathrm{d}\mathbf{v}$$

(cf. [5]). Using spherical coordinates in $\mathbb{R}^{N-m}$ (like [5], p. 427) we obtain the formula (8).

Suppose that $\boldsymbol{\varepsilon} \sim S_N(\mathbf{0}, \mathbf{I}, h)$. Denote $J := \{i_1, \ldots, i_m\}$. The conditional density of $\{\varepsilon_j : j \notin J\}$ given $\{\varepsilon_j : j \in J\}$ is evidently equal to

$$k_{N-m}\Big(\sum_{i \notin J} \varepsilon_i^2 \,\Big|\, \sum_{j \in J} \varepsilon_j^2\Big)$$

where

(11) $$k_{N-m}(t \mid u) := \frac{h(t+u)}{h_m(u)}.$$

Hence this density is spherically symmetrical.

Let $\boldsymbol{\varepsilon} \sim S_N(\mathbf{0}, \mathbf{I}, h)$. Then the probability density of the random variable $u := \|\boldsymbol{\varepsilon}\|^2$ is equal to

(12) $$\frac{\pi^{N/2}}{\Gamma\left(\frac{N}{2}\right)} u^{\frac{N}{2}-1} h(u)$$

(cf. [5]).

Evidently, if $\boldsymbol{\varepsilon} \sim S_N(\mathbf{0}, \mathbf{I}, h)$, then $\varepsilon_1, \ldots, \varepsilon_N$ are uncorrelated random variables. They are independent if and only if $f(\boldsymbol{\varepsilon})$ is the normal density (cf. [5] or [10], chpt. 3a.1).

We have a large choice for the function $h(t)$ in the expression (9). Some examples of $h(t)$ are (cf. [2]):

a)

$$h(t) = \alpha(2\pi)^{N/2} \int_0^\infty \exp\{-\tfrac{1}{2}\beta t u\}\, G(\mathrm{d}u)\,,$$

where G is a probability distribution on $\langle 0, \infty)$ and $\alpha > 0$, $\beta > 0$. The corresponding densities are mixed normal densities.

b)

$$h(t) = c t^{k-1} \exp\{-r t^{\lambda}\}$$

for some $c > 0$, $\lambda > 0$, $r > 0$ and k such that $2k + N > 2$ (the generalized gamma densities).

c)

$$h(t) = c\sqrt{(\pi/2)} \exp\{-\sqrt{(t)}/s\}\,,$$

where c, s are positive constants (the spherical Laplace density), etc.

3. THE GEOMETRY OF THE MODEL

The set

$$\mathscr{E} := \{\boldsymbol{\eta}(\boldsymbol{\theta}) : \boldsymbol{\theta} \in \boldsymbol{\Theta}\} \tag{13}$$

is the "expectation surface" of the nonlinear regression model (1). The point $\boldsymbol{\eta} = \boldsymbol{\eta}(\bar{\boldsymbol{\theta}})$ is a fixed point of $\mathscr{E}$. Take $r > 0$. Denote by

$$G_{\boldsymbol{\eta}}(r) := \{\mathbf{y} : \mathbf{y} \in \mathbb{R}^N, \|\mathbf{y} - \boldsymbol{\eta}\|_{\boldsymbol{\Sigma}} < r\} \tag{14}$$

a sphere centred at $\boldsymbol{\eta}$ (see Fig. 1). Further denote by $A_{\boldsymbol{\eta}}(r)$ a subset of the extended parameter space $\bar{\boldsymbol{\Theta}}$ defined by

$$A_{\boldsymbol{\eta}}(r) := \{\hat{\boldsymbol{\theta}}(\mathbf{y}) : \mathbf{y} \in G_{\boldsymbol{\eta}}(r)\}\,.$$

For every $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ denote by

$$\mathscr{N}(\boldsymbol{\theta}) := \left\{\mathbf{z} : \mathbf{z} \in \mathbb{R}^N, \mathbf{z}^{\mathrm{T}}\mathbf{V}^{-1}\frac{\partial\boldsymbol{\eta}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^{\mathrm{T}}} = \mathbf{0}\right\}$$

the subspace of $\mathbb{R}^N$ which is $\mathbf{V}$-orthogonal to the tangent plane to $\mathscr{E}$ (the later being generated by the vectors $\partial\boldsymbol{\eta}(\boldsymbol{\theta})/\partial\theta_1, \ldots, \partial\boldsymbol{\eta}(\boldsymbol{\theta})/\partial\theta_m$).

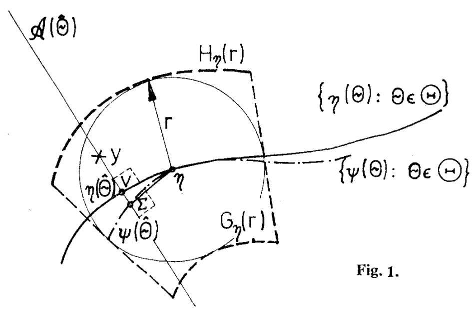

Fig. 1.

Denote by $\mathbf{w}_1(\boldsymbol{\theta}), \ldots, \mathbf{w}_{N-m}(\boldsymbol{\theta})$ a $\boldsymbol{\Sigma}$-orthogonal basis of $\mathcal{N}(\boldsymbol{\theta})$. It is $\mathbf{V}$-orthogonal to the tangent plane, i.e.

$$(15) \qquad \mathbf{w}_i^{\mathrm{T}}(\boldsymbol{\theta})\,\mathbf{V}^{-1}\,\frac{\partial \boldsymbol{\eta}(\boldsymbol{\theta})}{\partial \theta_j} = 0\,; \quad \begin{pmatrix} i = 1, \ldots, N-m \\ j = 1, \ldots, m \end{pmatrix}$$

$$\mathbf{w}_i^{\mathrm{T}}(\boldsymbol{\theta})\,\boldsymbol{\Sigma}^{-1}\mathbf{w}_j(\boldsymbol{\theta}) = 0 \quad \text{if} \quad i \neq j$$
$$= 1 \quad \text{if} \quad i = j$$

Evidently, the $\boldsymbol{\Sigma}$-orthogonal projector onto $\mathcal{N}(\boldsymbol{\theta})$ is equal to the matrix

$$\mathbf{R}^{\boldsymbol{\theta}} := \mathbf{W}(\boldsymbol{\theta})\,\mathbf{W}^{\mathrm{T}}(\boldsymbol{\theta})\,\boldsymbol{\Sigma}^{-1}$$

where $\mathbf{W}(\boldsymbol{\theta}) := (\mathbf{w}_1(\boldsymbol{\theta}), \ldots, \mathbf{w}_{N-m}(\boldsymbol{\theta}))$. Let us denote by

$$(16) \qquad \boldsymbol{\psi}(\boldsymbol{\theta}) := \boldsymbol{\eta}(\boldsymbol{\theta}) + \mathbf{R}^{\boldsymbol{\theta}}[\boldsymbol{\eta} - \boldsymbol{\eta}(\boldsymbol{\theta})]$$

the $\boldsymbol{\Sigma}$-orthogonal projection of the point $\boldsymbol{\eta}$ onto the set

$$\mathscr{A}(\boldsymbol{\theta}) := \mathcal{N}(\boldsymbol{\theta}) + \boldsymbol{\eta}(\boldsymbol{\theta})$$

(see Fig. 1 for $\mathbf{V} = \mathbf{I}$). We introduce the vector $\boldsymbol{\psi}(\boldsymbol{\theta})$ because $\boldsymbol{\psi}(\hat{\boldsymbol{\theta}})$ is equal to a conditional mean of $\mathbf{y}$ (see Section 4). We have

$$\boldsymbol{\psi}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\eta} = [\mathbf{I} - \mathbf{R}^{\hat{\boldsymbol{\theta}}}]\,(\boldsymbol{\eta}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\eta})\,,$$

and from (4) we obtain

$$\mathbf{y} - \boldsymbol{\eta}(\hat{\boldsymbol{\theta}}) \in \mathcal{N}(\hat{\boldsymbol{\theta}})\,.$$

Hence we have the Pythagorian relation

$$(17) \qquad \|\mathbf{y} - \boldsymbol{\eta}\|_{\boldsymbol{\Sigma}}^2 = \|\mathbf{y} - \boldsymbol{\psi}(\hat{\boldsymbol{\theta}})\|_{\boldsymbol{\Sigma}}^2 + \|\boldsymbol{\psi}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\eta}\|_{\boldsymbol{\Sigma}}^2\,.$$

Denote by

$$(18) \qquad H_{\boldsymbol{\eta}}(r) := \{\mathbf{y} : \mathbf{y} \in \mathbb{R}^N,\, \hat{\boldsymbol{\theta}}(\mathbf{y}) \in A_{\boldsymbol{\eta}}(r),\, \|\mathbf{y} - \boldsymbol{\psi}[\hat{\boldsymbol{\theta}}(\mathbf{y})]\|_{\boldsymbol{\Sigma}} < r\}$$

a "tube" in the sample space around the surface $\{\boldsymbol{\psi}(\boldsymbol{\theta}) : \boldsymbol{\theta} \in A_{\boldsymbol{\eta}}(r)\}$ (see Fig. 1). We have

$$(19) \qquad G_{\boldsymbol{\eta}}(r) \subset H_{\boldsymbol{\eta}}(r)\,.$$

In Section 4 we shall consider samples belonging to $H_{\boldsymbol{\eta}}(r)$, but only such that the corresponding L. S. estimates are not on the boundary of $\bar{\boldsymbol{\Theta}}$. Therefore we assume that:

A1:

$$A_{\boldsymbol{\eta}}(r) \subset \boldsymbol{\Theta}$$

(i.e. the point $\boldsymbol{\eta}$ is "sufficiently distant" from the boundary of $\bar{\boldsymbol{\Theta}}$).

To avoid complications with the nonidentifiability of the parameter $\boldsymbol{\theta}$ we shall suppose that

A2: The mapping $\boldsymbol{\theta} \in A_{\boldsymbol{\eta}}(r) \mapsto \boldsymbol{\eta}(\boldsymbol{\theta}) \in \mathscr{E}$ is one-to-one.

To avoid that the expectation surface $\mathscr{E}$ could overlap the neighbourhood of its subset $\{\boldsymbol{\eta}(\boldsymbol{\theta}): \boldsymbol{\theta} \in A_{\eta}(r)\}$, we require that r is such that

A3: If

i) $\mathbf{y} \in H_{\eta}(r)$

ii) $\boldsymbol{\theta}^*$ is a solution of (4)

iii) $\|\mathbf{y} - \boldsymbol{\eta}(\boldsymbol{\theta}^*)\|_{\boldsymbol{\Sigma}} < r$

then $\boldsymbol{\theta}^* \in A_{\eta}(r)$ and $\boldsymbol{\theta}^* = \hat{\boldsymbol{\theta}}(\mathbf{y})$.

Finally we shall suppose that

A4: The surface $\{\boldsymbol{\eta}(\boldsymbol{\theta}): \boldsymbol{\theta} \in A_{\eta}(r)\}$ has no centre of curvature which is a point of $H_{\eta}(r)$.

How to compute numerically curvatures of the expectation surface is explained in [1] and in the appendix of [7]. For a further use we present the definition of a geodesics on $\mathscr{E}$, like in [8].

By definition, a curve

$$\gamma: (-\delta, \delta) \mapsto \mathscr{E}$$

is a $\mathbf{V}$-geodesics on $\mathscr{E}$ through the point $\gamma(0) = \boldsymbol{\eta}(\bar{\boldsymbol{\theta}})$ if there is a twice continuously differentiable mapping

$$\boldsymbol{\varkappa}: (-\delta, \delta) \mapsto \boldsymbol{\Theta}$$

such that for every $t \in (-\delta, \delta)$

i) $\gamma(t) = \boldsymbol{\eta} \circ \boldsymbol{\varkappa}(t)$

ii) $\left\|\dfrac{\mathrm{d}\gamma(t)}{\mathrm{d}t}\right\|_{\mathbf{V}} = 1$

i.e. the parameter t is the length of the curve γ,

iii) $\dfrac{\mathrm{d}^2\boldsymbol{\gamma}^{\mathrm{T}}(t)}{\mathrm{d}t^2}\mathbf{V}^{-1}\left.\dfrac{\partial\boldsymbol{\eta}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^{\mathrm{T}}}\right|_{\boldsymbol{\theta}=\boldsymbol{\varkappa}(t)} = \mathbf{0}$

i.e. the "vector of curvature" $\mathrm{d}^2\boldsymbol{\gamma}^{\mathrm{T}}(t)/\mathrm{d}t^2$ is always orthogonal to $\mathscr{E}$.

The radius of curvature of $\gamma(t)$ at $t = 0$ is equal to

$$r_{\gamma}(0) := \left[\left\|\frac{\mathrm{d}^2\gamma(t)}{\mathrm{d}t^2}\right\|_{\mathbf{V}}^{-1}\right]_{t=0}$$

and it is the radius of a circle which is "as tangent as possible" to the curve $\gamma(t)$. According to iii) this circle with centre (= the centre of curvature)

$$\boldsymbol{\eta}(\boldsymbol{\theta}) + \left.\frac{\mathrm{d}^2\gamma(t)}{\mathrm{d}t^2}\right|_{t=0}$$

is also tangent to the expectation surface $\mathscr{E}$ at the point $\boldsymbol{\eta}(\boldsymbol{\theta})$, and its radius-vector is $\mathbf{V}$-orthogonal to the tangent plane. The centre of curvature of γ is considered as a centre of curvature of the surface $\mathscr{E}$ at the point $\boldsymbol{\eta}(\boldsymbol{\theta})$. Since there are many $\mathbf{V}$-

geodesics on $\mathscr{E}$ going through the same point $\boldsymbol{\eta}(\boldsymbol{\theta})$, we define the minimal radius of curvature

$$\varrho(\boldsymbol{\theta}) := \inf_{\gamma} r_{\gamma}(0) .$$

Instead of A4 we can assume equivalently

A4*: $$r < \varrho(\boldsymbol{\theta}) ; \quad (\boldsymbol{\theta} \in A_{\boldsymbol{\eta}}(r))$$

The assumptions A1 – A4 are slight modifications of the assumptions formulated in [7, 8]. A heuristic discussion is in [7].

The vector $\mathbf{y} - \boldsymbol{\psi}(\hat{\boldsymbol{\theta}})$ is $\mathbf{V}$-orthogonal to the tangent plane (Eqs. (4) and (16)), hence we can write

$$\mathbf{y} = \boldsymbol{\psi}(\hat{\boldsymbol{\theta}}) + \sum_{l=1}^{N-m} b_l \boldsymbol{w}_l(\hat{\boldsymbol{\theta}})$$

where

$$b_l := [\mathbf{y} - \boldsymbol{\psi}(\hat{\boldsymbol{\theta}})]^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \boldsymbol{w}_l(\hat{\boldsymbol{\theta}}) .$$

It follows that $\hat{\theta}_1, \ldots, \hat{\theta}_m, b_1, \ldots, b_{N-m}$ can be used as new coordinates of the point $\mathbf{y} \in H_{\boldsymbol{\eta}}(r)$. The corresponding coordinate transformation will be denoted by $\mathbf{g}(\hat{\boldsymbol{\theta}}, \boldsymbol{b})$, i.e.

$$\mathbf{g}(\hat{\boldsymbol{\theta}}, \boldsymbol{b}) := \boldsymbol{\psi}(\hat{\boldsymbol{\theta}}) + \sum_{l=1}^{N-m} b_l \boldsymbol{w}_l(\hat{\boldsymbol{\theta}}) \tag{20}$$

Its Jacobi matrix $\nabla \mathbf{g}$ is equal to

$$\begin{aligned} \nabla \mathbf{g}(\hat{\boldsymbol{\theta}}, \boldsymbol{b}) &:= \left(\frac{\partial \mathbf{g}}{\partial \hat{\boldsymbol{\theta}}^{\mathrm{T}}}, \frac{\partial \mathbf{g}}{\partial \boldsymbol{b}^{\mathrm{T}}} \right) \\ &= \left(\frac{\partial \mathbf{g}}{\partial \hat{\boldsymbol{\theta}}^{\mathrm{T}}}, \mathbf{W}(\hat{\boldsymbol{\theta}}) \right) \end{aligned}$$

Proposition 1. We have

$$|\det [\nabla \mathbf{g}(\hat{\boldsymbol{\theta}}, \boldsymbol{b})]| = \frac{\det [\mathbf{Q}(\hat{\boldsymbol{\theta}}, \bar{\boldsymbol{\theta}}) + \mathbf{D}(\boldsymbol{b}, \hat{\boldsymbol{\theta}})]}{\det^{1/2} \mathbf{B}(\hat{\boldsymbol{\theta}})} \det{}^{1/2} \boldsymbol{\Sigma} \tag{21}$$

where $\mathbf{Q}(\hat{\boldsymbol{\theta}}, \bar{\boldsymbol{\theta}})$ and $\mathbf{B}(\hat{\boldsymbol{\theta}})$ are defined in (6), and $\mathbf{D}(\boldsymbol{b}, \hat{\boldsymbol{\theta}})$ is an $m \times m$ matrix

$$\{\mathbf{D}(\boldsymbol{b}, \boldsymbol{\theta})\}_{ij} := -\sum_{l=1}^{N-m} b_l \boldsymbol{w}_l^{\mathrm{T}}(\boldsymbol{\theta}) \mathbf{V}^{-1} \frac{\partial^2 \boldsymbol{\eta}(\boldsymbol{\theta})}{\partial \theta_i \, \partial \theta_j} .$$

The proof is in the Appendix.

If we compare the right-hand side of Eq. (21) with the first term in the right-hand side of Eq. (5) we see that we omitted the matrix $\mathbf{D}(\boldsymbol{b}, \hat{\boldsymbol{\theta}})$ in the determinant in (5). To evaluate the influence of this omission we shall need the following Proposition 2.

Let us use the notation

$$\mathbf{e} := (b_1, \ldots, b_{N-m})^{\mathrm{T}} / \|\boldsymbol{b}\| .$$

We can write

(22) $$\mathbf{D}(\boldsymbol{b}, \hat{\boldsymbol{\theta}}) = \|\boldsymbol{b}\| \, \mathbf{D}(\mathbf{e}, \hat{\boldsymbol{\theta}}) .$$

For every $m \times m$ matrix $\mathbf{A}$, and $s \leqq m$, denote by $\mathbf{A}^{(s)}$ the matrix of all $s \times s$ minors of $\mathbf{A}$; hence $\operatorname{tr}[\mathbf{A}^{(s)}]$ is the sum of all $s \times s$ principal minors of $\mathbf{A}$ (cf. [3]).

Proposition 2. For every $\hat{\boldsymbol{\theta}} \in A_{\boldsymbol{\eta}}(r)$ we have

$$|\operatorname{tr}[\mathbf{D}(\mathbf{e}, \hat{\boldsymbol{\theta}}) \, \mathbf{Q}^{-1}(\hat{\boldsymbol{\theta}}, \bar{\boldsymbol{\theta}})]^{(s)}| \leqq \binom{m}{s} \frac{2^s}{r^s} .$$

Proposition 3. If the matrix $\mathbf{V}$ is dominating $\boldsymbol{\Sigma}$, i.e.. $\|\boldsymbol{a}\|_{\boldsymbol{\Sigma}} \geqq \|\boldsymbol{a}\|_{\mathbf{V}}$; ($\boldsymbol{a} \in \mathbb{R}^N$), then for every $\hat{\boldsymbol{\theta}} \in A_{\boldsymbol{\eta}}(r)$ the matrix $\mathbf{Q}(\hat{\boldsymbol{\theta}}, \bar{\boldsymbol{\theta}})$ is positive definite.

The proofs of both propositions are in the Appendix.

4. THE PROBABILITY DENSITY OF $\hat{\boldsymbol{\theta}}$

The probability density of $\mathbf{y}$ is given in Eq. (9). In the sequel we shall not take into account those samples $\mathbf{y}$ which belong to the set $\mathbb{R}^N - H_{\boldsymbol{\eta}}(r)$. From (12) and (19) it follows that the probability of this set is bounded above by the number

$$1 - \int_{G_{\boldsymbol{\eta}}(r)} f_Y(\mathbf{y}) \, d\mathbf{y} = \int_r^\infty \frac{\pi^{N/2}}{\Gamma(N/2)} u^{N/2-1} h(u) \, du .$$

For points inside the set $H_{\boldsymbol{\eta}}(r)$ we shall use the coordinate transformation (20), to obtain the joint density of $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{b}$:

$$p_{\boldsymbol{\eta}}(\hat{\boldsymbol{\theta}}, \boldsymbol{b}) := |\det[\nabla \mathbf{g}(\hat{\boldsymbol{\theta}}, \boldsymbol{b})]| \det^{-1/2}(\boldsymbol{\Sigma}) \, h(\|\boldsymbol{b}\|^2 + \|\boldsymbol{\psi}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\eta}\|^2_{\boldsymbol{\Sigma}})$$

where we used Eq. (17) and the equality $\|\boldsymbol{b}\|^2 = \|\mathbf{y} - \boldsymbol{\psi}(\hat{\boldsymbol{\theta}})\|^2_{\boldsymbol{\Sigma}}$. Denote $I(r) := := \langle -r, r \rangle^{N-m}$. The density of $\hat{\boldsymbol{\theta}}$ is the marginal density

(23) $$\tilde{p}_{\boldsymbol{\eta}}(\hat{\boldsymbol{\theta}}) := \int_{I(r)} p_{\boldsymbol{\eta}}(\hat{\boldsymbol{\theta}}, \boldsymbol{b}) \, d\boldsymbol{b} =$$

$$= \int_{I(r)} \frac{\det[\mathbf{Q}(\hat{\boldsymbol{\theta}}, \bar{\boldsymbol{\theta}}) + \mathbf{D}(\boldsymbol{b}, \hat{\boldsymbol{\theta}})]}{\det^{1/2} \mathbf{B}(\hat{\boldsymbol{\theta}})} h(\|\boldsymbol{b}\|^2 + \|\boldsymbol{\psi}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\eta}\|^2_{\boldsymbol{\Sigma}}) \, d\boldsymbol{b} \text{ (Proposition 1)} =$$

$$= q(\hat{\boldsymbol{\theta}} \mid \bar{\boldsymbol{\theta}}) \int_{I(r)} \det[\mathbf{I} + \mathbf{D}(\boldsymbol{b}, \hat{\boldsymbol{\theta}}) \, \mathbf{Q}^{-1}(\hat{\boldsymbol{\theta}}, \bar{\boldsymbol{\theta}})] \, k_{N-m}(\|\boldsymbol{b}\|^2 \mid \|\boldsymbol{\psi}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\eta}\|^2_{\boldsymbol{\Sigma}}) \, d\boldsymbol{b}$$

Here we used Eq. (11) and the equality

$$\boldsymbol{\psi}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\eta} = \mathbf{P}^{\hat{\boldsymbol{\theta}}}[\boldsymbol{\eta}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\eta}]$$

which follows from Eq. (16) and (A2).

Denote by $\mathsf{E}^*_{\hat{\boldsymbol{\theta}}}$ the (conditional) mean with respect to the density

$$\boldsymbol{b} \in I(r) \mapsto \varphi(\boldsymbol{b} \mid \hat{\boldsymbol{\theta}}) := k_{N-m}(\|\boldsymbol{b}\|^2 \mid \|\boldsymbol{\psi}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\eta}\|^2_{\boldsymbol{\Sigma}}) .$$

Instead of Eq. (23) we can write

(24) $$\tilde{p}_{\boldsymbol{\eta}}(\hat{\boldsymbol{\theta}}) = q(\hat{\boldsymbol{\theta}} \mid \bar{\boldsymbol{\theta}}) \, \mathsf{E}^*_{\hat{\boldsymbol{\theta}}}\{\det[\mathbf{I} + \mathbf{D}(\boldsymbol{b}, \hat{\boldsymbol{\theta}}) \, \mathbf{Q}^{-1}(\hat{\boldsymbol{\theta}}, \bar{\boldsymbol{\theta}})]\} .$$

From [4], III, §7 we obtain

(25) $$\det\left[\mathbf{I} + \mathbf{D}(\boldsymbol{b}, \hat{\boldsymbol{\theta}})\, \mathbf{Q}^{-1}(\hat{\boldsymbol{\theta}}, \bar{\boldsymbol{\theta}})\right] = 1 + \sum_{s=1}^{m} \operatorname{tr}\left[\mathbf{D}(\boldsymbol{b}, \hat{\boldsymbol{\theta}})\, \mathbf{Q}^{-1}(\hat{\boldsymbol{\theta}}, \bar{\boldsymbol{\theta}})\right]^{(s)}$$

According to the definition of $\mathbf{D}(\boldsymbol{b}, \hat{\boldsymbol{\theta}})$, each term in the right-hand side of Eq. (25) is a homogeneous polynomal in the variables $b_1, \ldots, b_{N-r}$. Consequently, if s is odd, then

$$\mathsf{E}^*_{\hat{\boldsymbol{\theta}}}\left[\mathbf{D}(\boldsymbol{b}, \hat{\boldsymbol{\theta}})\, \mathbf{Q}^{-1}(\hat{\boldsymbol{\theta}}, \bar{\boldsymbol{\theta}})\right]^{(s)} = \mathbf{0}\,,$$

because $\varphi(\boldsymbol{b} \mid \hat{\boldsymbol{\theta}})$ is a spherically symmetrical density. It follows that

(26) $$\mathsf{E}^*_{\hat{\boldsymbol{\theta}}}\{\det\left[\mathbf{I} + \mathbf{D}(\boldsymbol{b}, \hat{\boldsymbol{\theta}})\, \mathbf{Q}^{-1}(\hat{\boldsymbol{\theta}}, \bar{\boldsymbol{\theta}})\right]\} \leqq$$
$$\leqq 1 + \sum_{s=1}^{\mathrm{INT}(m/2)} \mathsf{E}^*_{\hat{\boldsymbol{\theta}}}\{\left|\operatorname{tr}\left[\mathbf{D}(\boldsymbol{b}, \hat{\boldsymbol{\theta}})\, \mathbf{Q}^{-1}(\hat{\boldsymbol{\theta}}, \bar{\boldsymbol{\theta}})\right]^{(2s)}\right|\} \leqq$$
$$\leqq 1 + \sum_{s=1}^{\mathrm{INT}(m/2)} \mathsf{E}^*_{\hat{\boldsymbol{\theta}}}(\|\boldsymbol{b}\|^{2s}) \binom{m}{2s} \left(\frac{2}{r}\right)^{2s}$$

(Eq. (22) and Proposition 2.).

Similarly we obtain

(27) $$\mathsf{E}^*_{\hat{\boldsymbol{\theta}}}\{\det\left[\mathbf{I} + \mathbf{D}(\boldsymbol{b}, \hat{\boldsymbol{\theta}})\, \mathbf{Q}^{-1}(\hat{\boldsymbol{\theta}}, \bar{\boldsymbol{\theta}})\right]\} \geqq$$
$$\geqq 1 - \sum_{s=1}^{\mathrm{INT}(m/2)} \mathsf{E}^*_{\hat{\boldsymbol{\theta}}}(\|\boldsymbol{b}\|^{2s}) \binom{m}{2s} \left(\frac{2}{r}\right)^{2s}.$$

Further, we have from Eqs. (8) and (12)

$$\mathsf{E}^*_{\hat{\boldsymbol{\theta}}}(\|\boldsymbol{b}\|^{2s}) \leqq \int_{\|\boldsymbol{b}\|^2 \leqq (N-m) r^2} \|\boldsymbol{b}\|^{2s} \frac{h(\|\boldsymbol{b}\|^2 + \|\boldsymbol{\psi}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\eta}\|^2_{\boldsymbol{\Sigma}})}{h_m(\|\boldsymbol{\psi}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\eta}\|^2_{\boldsymbol{\Sigma}})}\, \mathrm{d}\boldsymbol{b} =$$
$$= \frac{\int_0^{(N-m)r^2} u^s u^{(N-m)/2-1}\, h(u + \|\boldsymbol{\psi}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\eta}\|^2_{\boldsymbol{\Sigma}})\, \mathrm{d}u}{\int_0^\infty u^{(N-m)/2-1}\, h(u + \|\boldsymbol{\psi}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\eta}\|^2_{\boldsymbol{\Sigma}})\, \mathrm{d}u}.$$

Consequently, if h is a nonincreasing function, then from $\|\boldsymbol{\psi}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\eta}\|_{\boldsymbol{\Sigma}} < r$ we obtain

(28) $$\mathsf{E}^*_{\hat{\boldsymbol{\theta}}}(\|\boldsymbol{b}\|^{2s}) \leqq \frac{\int_0^{(N-m)r^2} u^{(N-m)/2+s-1}\, h(u)\, \mathrm{d}u}{\int_0^{(N-m)r^2} u^{(N-m)/2-1}\, h(r^2 + u)\, \mathrm{d}u}.$$

From Eqs (24)–(28) follows the proof of the following theorem.

Theorem. If $h: \langle 0, \infty) \mapsto \langle 0, \infty)$ is non-increasing then

$$\frac{\tilde{p}_{\boldsymbol{\eta}}(\hat{\boldsymbol{\theta}}) - q(\hat{\boldsymbol{\theta}} \mid \bar{\boldsymbol{\theta}})}{q(\hat{\boldsymbol{\theta}} \mid \bar{\boldsymbol{\theta}})} \leqq \sum_{s=1}^{\mathrm{INT}(m/2)} \binom{m}{2s} \left(\frac{2}{r}\right)^{2s} \frac{\int_0^{(N-m)r^2} u^{(N-m)/2+s-1}\, h(u)\, \mathrm{d}u}{\int_0^{(N-m)r^2} u^{(N-m)/2-1}\, h(r^2 + u)\, \mathrm{d}u}$$

APPENDIX

Proof of Proposition 1. We shall write $\boldsymbol{\theta}$ instead of $\hat{\boldsymbol{\theta}}$. We have

$$\text{(A1)} \qquad \frac{\det^2[\nabla \mathbf{g}(\boldsymbol{\theta}, \boldsymbol{b})]}{\det \boldsymbol{\Sigma}} = \det \begin{pmatrix} \frac{\partial \mathbf{g}^{\mathrm{T}}}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}^{-1} \frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}^{\mathrm{T}}}, & \frac{\partial \mathbf{g}^{\mathrm{T}}}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}^{-1} \mathbf{W} \\ \mathbf{W}^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}^{\mathrm{T}}}, & \mathbf{I} \end{pmatrix} =$$

$$= \det \left(\frac{\partial \mathbf{g}^{\mathrm{T}}}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}^{-1} [\mathbf{I} - \mathbf{W}\mathbf{W}^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}] \frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}^{\mathrm{T}}} \right) \quad ([4], \text{II}, \S 5)$$

$$= \det \left(\frac{\partial \mathbf{g}^{\mathrm{T}}}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma}^{-1} [\mathbf{I} - \mathbf{R}^{\boldsymbol{\theta}}] \frac{\partial \mathbf{g}}{\partial \boldsymbol{\theta}^{\mathrm{T}}} \right).$$

From the equation

$$\mathbf{R}^{\boldsymbol{\theta}} \boldsymbol{\Sigma} \mathbf{V}^{-1} \frac{\partial \boldsymbol{\eta}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{\mathrm{T}}} = \mathbf{W}(\boldsymbol{\theta}) \mathbf{W}^{\mathrm{T}}(\boldsymbol{\theta}) \mathbf{V}^{-1} \frac{\partial \boldsymbol{\eta}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{\mathrm{T}}} = \mathbf{0}$$

we see that the linearly independent vectors

$$\mathbf{t}_i := \boldsymbol{\Sigma} \mathbf{V}^{-1} \frac{\partial \boldsymbol{\eta}(\boldsymbol{\theta})}{\partial \theta_i}; \quad (i = 1, \ldots, m)$$

span the linear space $\{\mathbf{z}: \mathbf{z} \in \mathbb{R}^N, (\mathbf{I} - \mathbf{R}^{\boldsymbol{\theta}})\, \mathbf{z} = \mathbf{z}\}$. Hence the $\boldsymbol{\Sigma}$-orthogonal projector onto this space is equal to

$$\mathbf{I} - \mathbf{R}^{\boldsymbol{\theta}} = \sum_{i,j=1}^{m} \mathbf{t}_i \{\mathbf{T}^{-1}\}_{ij}\, \mathbf{t}_j^{\mathrm{T}} \boldsymbol{\Sigma}^{-1}$$

where $\{\mathbf{T}\}_{ij} := \mathbf{t}_i^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \mathbf{t}_j$. It is easy to verify that $\mathbf{T} = \mathbf{B}(\boldsymbol{\theta})$, and that

$$\text{(A2)} \qquad \mathbf{I} - \mathbf{R}^{\boldsymbol{\theta}} = \mathbf{P}^{\boldsymbol{\theta}},$$

where $\mathbf{P}^{\boldsymbol{\theta}}$ is defined in Eq. (7). Putting the expression for $\mathbf{P}^{\boldsymbol{\theta}}$ into (A1) we obtain

$$\frac{\det^2[\nabla \mathbf{g}(\boldsymbol{\theta}, \boldsymbol{b})]}{\det \boldsymbol{\Sigma}} = \frac{\det^2 \left(\frac{\partial \mathbf{g}^{\mathrm{T}}}{\partial \boldsymbol{\theta}} \mathbf{V}^{-1} \frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\theta}^{\mathrm{T}}} \right)}{\det [\mathbf{B}(\boldsymbol{\theta})]} =$$

$$= \frac{\det^2 \left[\frac{\partial \boldsymbol{\psi}^{\mathrm{T}}}{\partial \boldsymbol{\theta}} \mathbf{V}^{-1} \frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\theta}^{\mathrm{T}}} + \sum_l b_l \frac{\partial \mathbf{w}_l^{\mathrm{T}}}{\partial \boldsymbol{\theta}} \mathbf{V}^{-1} \frac{\partial \boldsymbol{\eta}}{\partial \boldsymbol{\theta}^{\mathrm{T}}} \right]}{\det [\mathbf{B}(\boldsymbol{\theta})]}.$$

From

$$\boldsymbol{\psi}(\boldsymbol{\theta}) - \boldsymbol{\eta}(\boldsymbol{\theta}) = \mathbf{R}^{\boldsymbol{\theta}} [\boldsymbol{\eta} - \boldsymbol{\eta}(\boldsymbol{\theta})] \in \mathcal{N}(\boldsymbol{\theta})$$

we obtain that

$$\text{(A3)} \qquad [\boldsymbol{\psi}(\boldsymbol{\theta}) - \boldsymbol{\eta}(\boldsymbol{\theta})]^{\mathrm{T}} \mathbf{V}^{-1} \frac{\partial \boldsymbol{\eta}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{\mathrm{T}}} = \mathbf{0}.$$

We differentiate this equality, and obtain

$$\text{(A4)} \qquad \frac{\partial \boldsymbol{\psi}^{\mathrm{T}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{V}^{-1} \frac{\partial \boldsymbol{\eta}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{\mathrm{T}}} = \mathbf{M}(\boldsymbol{\theta}) + [\boldsymbol{\eta}(\boldsymbol{\theta}) - \boldsymbol{\psi}(\boldsymbol{\theta})]^{\mathrm{T}} \mathbf{V}^{-1} \frac{\partial^2 \boldsymbol{\eta}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{\mathrm{T}}} = \mathbf{Q}(\boldsymbol{\theta}, \bar{\boldsymbol{\theta}}).$$

Further, differentiating the first equations in (15) we obtain

$$\mathbf{D}(\boldsymbol{b}, \boldsymbol{\theta}) = \sum_i b_i \frac{\partial \boldsymbol{w}_i^{\mathrm{T}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \mathbf{V}^{-1} \frac{\partial \boldsymbol{\eta}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{\mathrm{T}}}.$$

Finally, from (3) it follows that the matrix $\partial^2/\partial \boldsymbol{\theta}\, \partial \boldsymbol{\theta}^{\mathrm{T}} \{\frac{1}{2} \|\boldsymbol{\eta}(\boldsymbol{\theta}) - \boldsymbol{y}\|_{\boldsymbol{\Sigma}}^2\}_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$ is p.d., and we can verify that it is equal to $\mathbf{Q}(\hat{\boldsymbol{\theta}}, \bar{\boldsymbol{\theta}}) + \mathbf{D}(\boldsymbol{b}, \hat{\boldsymbol{\theta}})$ when putting $\boldsymbol{y} = \boldsymbol{g}(\hat{\boldsymbol{\theta}}, \boldsymbol{b})$. □

The matrix $\mathbf{M}(\boldsymbol{\theta})$ is positive definite. Therefore, there is a nonsingular matrix $\mathbf{U}$ such that

$$\mathbf{U}^{\mathrm{T}} \mathbf{M}(\boldsymbol{\theta}) \mathbf{U} = \mathbf{I}.$$

Denote

$$\mathbf{D}^*(\mathbf{e}) := \mathbf{D}^*(\mathbf{e}, \boldsymbol{\theta}) := \mathbf{U}^{\mathrm{T}} \mathbf{D}(\mathbf{e}, \boldsymbol{\theta}) \mathbf{U}$$
$$\mathbf{Q}^* := \mathbf{Q}^*(\boldsymbol{\theta}) := \mathbf{U}^{\mathrm{T}} \mathbf{Q} \mathbf{U}.$$

For any eigenvalue λ of the matrix $\mathbf{D}(\mathbf{e}, \boldsymbol{\theta})$ we have the inequality

$$\text{(A5)} \qquad |\lambda| \leqq \frac{1}{\varrho(\boldsymbol{\theta})}$$

(cf. [7], Proposition 2).

Proof of Proposition 2. For any matrices $\mathbf{A}$, $\mathbf{B}$ we have (cf. [3], theorem 6.13)

$$\mathbf{A}^{(s)} \mathbf{B}^{(s)} = (\mathbf{AB})^{(s)}.$$

Hence

$$\text{(A6)} \qquad \operatorname{tr}[\mathbf{D}^*(\mathbf{e})\, \mathbf{Q}^{*-1}]^{(s)} = \operatorname{tr}\{(\mathbf{U}^{\mathrm{T}})^{(s)} [\mathbf{D}(\mathbf{e})\, \mathbf{Q}^{-1}]^{(s)} (\mathbf{U}^{\mathrm{T}(-1)})^{(s)}\}$$
$$= \operatorname{tr}[\mathbf{D}(\mathbf{e})\, \mathbf{Q}^{-1}]^{(s)}$$

Denote by $\mathbf{C} := (\mathbf{c}^{(1)}, \ldots, \mathbf{c}^{(m)})$ and by $\boldsymbol{\Lambda} := \operatorname{diag}(\lambda_1, \ldots, \lambda_m)$ the matrices of the orthonormal eigenvectors and of the eigenvalues of $\mathbf{D}^*(\mathbf{e})$. From $\mathbf{D}^*(\mathbf{e}) = \mathbf{C} \boldsymbol{\Lambda} \mathbf{C}^{\mathrm{T}}$ we obtain

$$\operatorname{tr}[\mathbf{D}^*(\mathbf{e})\, \mathbf{Q}^{*-1}]^{(s)} = \operatorname{tr}[\boldsymbol{\Lambda}^{(s)} (\mathbf{C}^{\mathrm{T}})^{(s)} (\mathbf{Q}^{*-1})^{(s)} \mathbf{C}^{(s)}].$$

The matrix $\boldsymbol{\Lambda}^{(s)}$ is diagonal, having diagonal entries of the form $\lambda_{i_1}, \ldots, \lambda_{i_s}$; $(i_1 < \ldots \ldots < i_s)$. Hence from (A5) we obtain

$$\text{(A7)} \qquad |\operatorname{tr}[\mathbf{D}^*(\mathbf{e})\, \mathbf{Q}^{*-1}]^{(s)}| \leqq [\varrho(\boldsymbol{\theta})]^{-s} \operatorname{tr}[(\mathbf{C}^{\mathrm{T}})^{(s)} (\mathbf{Q}^{*-1})^{(s)} \mathbf{C}^{(s)}] =$$
$$= [\varrho(\boldsymbol{\theta})]^{-s} \operatorname{tr}(\mathbf{Q}^{*-1})^{(s)}$$

since $\mathbf{C}^{(s)} (\mathbf{C}^{\mathrm{T}})^{(s)} = (\mathbf{C}\mathbf{C}^{\mathrm{T}})^{(s)} = \mathbf{I}^{(s)} = \mathbf{I}$.

From (A3) we obtain

$$\mathbf{Q} = \mathbf{I} + \mathbf{U}^{\mathrm{T}} [\boldsymbol{\eta}(\boldsymbol{\theta}) - \boldsymbol{\psi}(\boldsymbol{\theta})]^{\mathrm{T}} \mathbf{V}^{-1} \frac{\partial^2 \boldsymbol{\eta}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\, \partial \boldsymbol{\theta}^{\mathrm{T}}} \mathbf{U}.$$

According to Eq. (A3) we can write

$$\boldsymbol{\eta}(\boldsymbol{\theta}) - \boldsymbol{\psi}(\boldsymbol{\theta}) = -\sum_{i=1}^{N-m} d_i \, \boldsymbol{w}_i(\boldsymbol{\theta}),$$

where $d_i := [\boldsymbol{\psi}(\boldsymbol{\theta}) - \boldsymbol{\eta}(\boldsymbol{\theta})]^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} \boldsymbol{w}_i(\boldsymbol{\theta})$. Hence

$$[\boldsymbol{\eta}(\boldsymbol{\theta}) - \boldsymbol{\psi}(\boldsymbol{\theta})]^{\mathrm{T}} \mathbf{V}^{-1} \frac{\partial^2 \boldsymbol{\eta}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \, \partial \boldsymbol{\theta}^{\mathrm{T}}} = \mathbf{D}(\boldsymbol{d}, \boldsymbol{\theta}).$$

Thus

$$\mathbf{Q}^* = \mathbf{I} + \|\boldsymbol{\psi}(\boldsymbol{\theta}) - \boldsymbol{\eta}(\boldsymbol{\theta})\|_{\boldsymbol{\Sigma}} \, \mathbf{D}^*(\boldsymbol{f})$$

where $\boldsymbol{f} := \boldsymbol{d}/\|\boldsymbol{d}\|$.

Using once more the inequality (A4) we obtain that the eigenvalues $\mu_1, \ldots, \mu_m$ of the matrix $\mathbf{Q}^*$ are bounded according to the inequalities

$$1 - \|\boldsymbol{\psi}(\boldsymbol{\theta}) - \boldsymbol{\eta}(\boldsymbol{\theta})\|_{\boldsymbol{\Sigma}} \, \varrho^{-1}(\boldsymbol{\theta}) \leqq \mu_i \leqq 1 + \|\boldsymbol{\psi}(\boldsymbol{\theta}) - \boldsymbol{\eta}(\boldsymbol{\theta})\|_{\boldsymbol{\Sigma}} \, \varrho^{-1}(\boldsymbol{\theta}).$$

Denote by $\mathbf{Z} := (\mathbf{z}_1, \ldots, \mathbf{z}_m)$ the matrix of the orthonormal eigenvectors of $\mathbf{Q}^*$. We have

(A8) $$\operatorname{tr}(\mathbf{Q}^{*-1})^{(s)} = \operatorname{tr}[(\mathbf{Z}^{-1}(\mathbf{Z}^{\mathrm{T}})^{-1})^{(s)} (\mathbf{Q}^{*-1})^{(s)}] = \operatorname{tr}[(\mathbf{Z}^{\mathrm{T}}\mathbf{Q}^*\mathbf{Z})^{-1}]^{(s)} =$$
$$= \sum_{i_1 < \ldots < i_s} \mu_{i_1}^{-1} \ldots \mu_{i_s}^{-1} \leqq \binom{m}{s} \left[\frac{\varrho(\boldsymbol{\theta})}{\varrho(\boldsymbol{\theta}) - \|\boldsymbol{\psi}(\boldsymbol{\theta}) - \boldsymbol{\eta}(\boldsymbol{\theta})\|_{\boldsymbol{\Sigma}}}\right]^s.$$

From (A6)–(A8) we have

$$|\operatorname{tr}[\mathbf{D}(\mathbf{e}, \boldsymbol{\theta}) \, \mathbf{Q}^{-1}(\boldsymbol{\theta}, \bar{\boldsymbol{\theta}})]^{(s)}| \leqq \binom{m}{s} [\varrho(\boldsymbol{\theta}) - \|\boldsymbol{\psi}(\boldsymbol{\theta}) - \boldsymbol{\eta}(\boldsymbol{\theta})\|_{\boldsymbol{\Sigma}}]^{-s}.$$

We obtain the required inequality from $[\boldsymbol{\psi}(\boldsymbol{\theta}) - \boldsymbol{\eta}(\boldsymbol{\theta})\|_{\boldsymbol{\Sigma}} \leqq \varrho(\boldsymbol{\theta})/2$ which follows from the assumption A4. □

Proof of Proposition 3. Take $\boldsymbol{\theta} \in A_{\boldsymbol{\eta}}(r)$, It is sufficient to show that for every geodesics $\gamma = \boldsymbol{\eta} \circ \boldsymbol{\varkappa}$ going through the point $\boldsymbol{\theta}$ the inequality

$$\frac{\mathrm{d}\boldsymbol{\varkappa}^{\mathrm{T}}(0)}{\mathrm{d}t} \mathbf{Q}(\boldsymbol{\theta}, \bar{\boldsymbol{\theta}}) \frac{\mathrm{d}\boldsymbol{\varkappa}(0)}{\mathrm{d}t} > 0$$

holds. From Eqs. (A4) and (16) we obtain

$$\frac{\mathrm{d}\boldsymbol{\varkappa}^{\mathrm{T}}}{\mathrm{d}t} \mathbf{Q}(\boldsymbol{\theta}, \bar{\boldsymbol{\theta}}) \frac{\mathrm{d}\boldsymbol{\varkappa}}{\mathrm{d}t} = \frac{\mathrm{d}\gamma^{\mathrm{T}}(0)}{\mathrm{d}t} \mathbf{V}^{-1} \frac{\mathrm{d}\gamma(0)}{\mathrm{d}t} + \frac{\mathrm{d}\boldsymbol{\varkappa}^{\mathrm{T}}(0)}{\mathrm{d}t} \left\{[\mathbf{R}^{\boldsymbol{\theta}}(\boldsymbol{\eta}(\boldsymbol{\theta}) - \boldsymbol{\eta})]^{\mathrm{T}} \mathbf{V}^{-1} \frac{\partial^2 \boldsymbol{\eta}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \, \partial \boldsymbol{\theta}^{\mathrm{T}}}\right\} \frac{\mathrm{d}\boldsymbol{\varkappa}(0)}{\mathrm{d}t}.$$

Hence from the definition of the $\mathbf{V}$-geodesics γ (Section 3) we have

$$\frac{\mathrm{d}\boldsymbol{\varkappa}^{\mathrm{T}}}{\mathrm{d}t} \mathbf{Q}(\boldsymbol{\theta}, \bar{\boldsymbol{\theta}}) \frac{\mathrm{d}\boldsymbol{\varkappa}}{\mathrm{d}t} = 1 + [\mathbf{R}^{\boldsymbol{\theta}}(\boldsymbol{\eta}(\boldsymbol{\theta}) - \eta)]^{\mathrm{T}} \mathbf{V}^{-1} \frac{\mathrm{d}^2\gamma(0)}{\mathrm{d}t^2} =$$

$$= 1 - [\boldsymbol{\psi}(\boldsymbol{\theta}) - \boldsymbol{\eta}(\boldsymbol{\theta})]^{\mathrm{T}} \mathbf{V}^{-1} \frac{\mathrm{d}^2\gamma(0)}{\mathrm{d}t^2}.$$

Therefore, from the Schwarz inequality and from the definition of $r_\gamma(0)$ (Section 3) we obtain that it is sufficient to prove that

$$\|\psi(\theta) - \eta(\theta)\|_V < r_\gamma(0)\,.$$

Since $\theta \in A_\eta(r)$, there is a point $\mathbf{y} \in G_\eta(r)$, i.e.

$$\|\mathbf{y} - \eta\|_\Sigma < r < r_\gamma(0)\,,$$

such that $\mathbf{y} \in \mathscr{A}(\theta)$ (see the definition of $A_\eta(r)$). Consequently

$$\|\psi(\theta) - \eta\|_\Sigma = \|\mathbf{P}^\theta(\mathbf{y} - \eta)\|_\Sigma < \|\mathbf{y} - \eta\|_\Sigma < r_\gamma(0)\,.$$

It follows that $\psi(\theta) \in G_\eta(r)$. Evidently $\psi(\theta) \in \mathscr{A}(\theta)$. Consequently, according to the property A3, θ solves Eq. (3) for $\mathbf{y} = \psi(\theta)$. It follows that $\|\psi(\theta) - \eta(\theta)\|_V \leqq \|\psi(\theta) - \eta\|_V \leqq \|\psi(\theta) - \eta\|_\Sigma < r_\gamma(0)$ since $\mathbf{V}$ is dominating Σ. $\square$

(Received March 10, 1988.)

REFERENCES

[1] D. M. Bates and D. G. Watts: Relative curvature measures of nonlinearity. J. R. Statist. Soc. B *42* (1980), 1–25.

[2] T. Cacoullos: On minimum-distance location discrimination for isotropic distributions. In: Proc. DIANA II Conf. on Discriminant Analysis, Cluster Analysis. Mathematical Inst., Czech. Acad. Sciences, Prague 1987, 1–16.

[3] M. Fiedler: Special Matrices and Their Use in Numerical Mathematics (in Czech). SNTL, Prague 1981.

[4] F. R. Gantmacher: Matrix Theory (in Russian). Nauka, Moscow 1966.

[5] D. Kelker: Distribution theory of spherical distributions and a location-scale parameter generalization. Sankhya *32A* (1970), 419–430.

[6] Yu. G. Kuritsin: On the least squares method for elliptically countered distributions (in Russian). Teor. Veroyatnost. i Primenen. *31* (1986), 834–838.

[7] A. Pázman: Probability distribution of the multivariate nonlinear least squares estimates. Kybernetika 20 (1984), 209–230.

[8] A. Pázman: On formulas for the distribution of nonlinear L.S. estimates. Statistics *18* (1987), 3–15.

[9] A. Pázman: On information matrices in nonlinear experimental design. J. Statist. Plann. Interference (in print).

[10] C. R. Rao: Linear Statistical Inference and Its Applications. Second edition. J. Wiley, New York 1973.

RNDr. Andrej Pázman, DrSc., Matematický ústav SAV (Mathematical Institute — Slovak Academy of Sciences), Obrancov mieru 49, 814 73 Bratislava. Czechoslovakia.